Meaningful Integration of Data, Analytics and Services

Grant Agreement No. 727721

Contract Duration: 40 months (1st November 2016 – 29th February 2020)

This project is funded by
The European Union

H2020-SC1-2016-CNECT
SC1-PM-18-2016 - Big Data Supporting Public Health Policies

Deliverable D5.2

Visual analytics tool(s) concept V2

Circulation:	PU (Public)
Nature:	R (Document/Report)
Version #:	1.0
Issue Date:	11/01/2019
Responsible Partner(s):	Teknologian tutkimuskeskus VTT Oy (VTT)
Author(s):	Juha Pajula, Adil Umer, Markus Ylikerälä; VTT, Gorka Epelde; Roberto Álvarez; VICOM, Joao Pita Costa, Flavio Fuart, Dunja Mladenic, Luka Stopar, Marko Grobelnik; QUIN, Jason Lloyd; IBM, Julius Gomes, Elina Hyrkäs, Jarmo Pääkkönen; UOULU, Oihana Belar, Roberto Bilbao; BIOEF, Paula Kavanagh, Chris Plunkett; HSE, Susan Campbell, Conor Fullerton; BSO (HSCNI), Scott Fischhaber; Analytics Eng
Status:	Final
Reviewed on:	10/01/2019

Grant Agreement No: 727721

Reviewed by:	Executive Board
Contractual Date of Delivery:	31/10/2018

Executive Board Document Sign Off

Role	Partner	Signature	Date
Project Coordinator	Ulster	Michaela Black	29/10/2018
WP1 Lead	Ulster	Michaela Black	29/10/2018
WP2 Lead	SET	Paul Carlin	31/10/2018
WP3 Lead	VICOM	Gorka Epelde	29/10/2018
WP4 Lead	KU Leuven	Gorana Nikolic	31/10/2018
WP5 Lead	VTT	Juha Pajula	10/01/2019
WP6 Lead	DCU	Regina Connolly	02/11/2018
WP7 Lead	Ulster	Jonathan Wallace	22/10/2018
WP8 Lead	Ulster	Michaela Black	29/10/2018
Scientific-Technical Manager	Analytics Eng	Scott Fischaber	30/10/2018

Abstract

This deliverable is the second version of Visual Analytics Tools Concept and it describes the updates and new approaches decided for the system and pilots after V1 implementation. The content includes descriptions of the four MIDAS pilot sites including stakeholders, scenarios and user stories in relation to user experience testing of the MIDAS Platform. It reports the implementation plan and agreed agile approach with a schedule for the testing of the MIDAS system. The first part of the document describes the Pilot details in respect to testing and development of the user interface. The latter part of document describes the agile development approach and selected technological approaches to implement the needed systems at version 2 of the MIDAS Platform.

Copyright

The MIDAS project is funded under the EC Horizon 2020 SC1- PMF-18 Big Data Supporting Public Health Policies

Document History

Version	Issue Date	Stage	Content and Changes
0.1	12.07.2018	Draft	Initial draft based on D5.1
0.1.1	11.09.2018	Draft	Updated structure to cover the agile approach
0.9	22.10.2018	Draft	Draft ready for Executive Board review
1.0	11.01.2019	Final	Final version after EB review corrections

Statement of Originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Executive Summary

Work Package:	Big Data Visualisation for Public Health Decision Making
Work Package leader:	VTT – Teknologian tutkimuskeskus VTT Oy
Task:	T 5.1 – Design of MIDAS Dashboard for Actionable Information Presentation
Task leader:	VTT – Teknologian tutkimuskeskus VTT Oy

This deliverable is the second version of Visual Analytics Tools Concept and it describes the updates and new approaches decided for the system and pilots after V1 implementation. The scenarios, use cases and stakeholder descriptions of the pilots are updated to reflect the agile development approach of the system with the updated knowledge from cases. The updated agile approach is used to ensure that the developed system answers to the requirements of the Policy Board as well as possible. Each pilot has now described the aimed test users via dedicated personas and the user stories, and use cases are focused to support the UX testing of the system within the pilotwise topics.

The selected technology and current state of MIDAS Dashboard development are reported together with implementation plans of version V2 of the MIDAS Dashboard together with whole MIDAS Platform. The main difference here in comparison to D5.1 are the continuous end user updates through the Agile testing protocol. The agreed schedule is reported on the deliverable from this fall to end of the project.

Some of the components and technologies were found challenging to implement on this phase and for this reason some of these were postponed to implementation V2 and even to V3 of the MIDAS Platform. These are for example system dynamics model runtime implementation to the MIDAS Dashboard and StreamStory technology. Also, more advanced authentication approaches are studied but not yet used in MIDAS Dashboard.

Table of Contents

1 Introduction

This document describes the current state of MIDAS Dashboard use cases, implementation and planned test schedule. It also describes the identified functionality of the MIDAS Dashboard which is needed to support the end users and plans how and in which order they are implemented for the version 2 of the MIDAS Platform. At the end of the document an overview of the final development phase to deliver version 3 of MIDAS Dashboard is introduced.

The document is divided into three main sections: Pilot Requirements, MIDAS Dashboard Architecture and Implementation Plan. The Pilot section describes the base for agile testing and developing approach used to develop the V2 of MIDAS Dashboard, the pilot use cases, test users and scenarios that test protocols are based on. The second part describes the current components and technology which is used to develop MIDAS Dashboard and the final section reports the agreed plans and schedules for development processes of V2 and V3 of MIDAS Dashboard. Deliverable D5.7 will report the work done within UX testing with reference to the setup described in this deliverable.

1.1 The Pilotwise KPIs

Here we have repeated the KPIs defined in D5.1 to remind the reader about the aims of the system on each MIDAS Pilot. These KPIs will be used as measures in UX testing on the corresponding pilot sites to identify how well the requirements of the end users are fulfilled. Each pilot has a different approach to the study setup which also makes the KPI definitions different between MIDAS Pilots.

1.1.1 Basque Country

- The MIDAS platform should enable a user to understand and identify the etiology of childhood obesity in the Basque Country and potential areas of intervention
- MIDAS platform should enable the Basque pilot:
    - To provide and clearly identify both crude and adjusted rates of target variables.
    - To provide information at different granularity levels, adequate for each stakeholder.
    - To cover time, place, and person analysis for epidemiological studies

1.1.2 Finland

Research question 1 (regional): How to use Midas platform and rich data from various sources in social and healthcare units in the regional stakeholder organizations to support preventive policy making?

- MIDAS should enable us to understand

  - what is the current practice for preventive policy making
  - to what extent data, digital services, analytics, visualizations are currently being used for aiding decision making?
  - what are the gaps in the policy makers decisions, how are the gaps affecting to their work?
  - What sort of data, analytics, visualizations could help decision makers in the preventive decision making?
  - Which data sources can be combined in order to create geographic and demographic analytics?

Research question 2 (national): What kind of regulations, enablers, metrics and should be enforced;

- How to identify correlations between the different parameters related to the policy maker decisions escalating on various government sectors;
- How can we simulate the different systemic correlations and impact factors to the policy makers to help their decisions?

1.1.3 Northern Ireland

- The MIDAS platform should enable us to carry out a longitudinal analysis and track a cohort of Looked After Children as they move in and out of care, use a variety of health services, and look at patterns of behaviours and changes over time. We are currently unable to do this analysis.
- The MIDAS platform should enable analysis of the available datasets to identify effective preventative measures.
- The MIDAS platform should enable the use of visual analytics to assess the impact of changing variables and indicators on policy decisions
- The MIDAS platform should enable the harmonisation and integration of multiple data sources

1.1.4 Republic of Ireland

- KPI 1: The provision of an adoptable intelligent analytics platform for stakeholder use
- KPI 2: Identification of the cohort of persons with diabetes in the Republic of Ireland
- KPI 3: Mapping of the care journey for people with diabetes through the health service
- KPI 4: Determining of information that can be shared to facilitate better use of resources and services for people with diabetes
- KPI 5: Determining the best fit for diabetes services and their locations based on population based geographical needs
- KPI 6: Improving prescribing and reducing hospital admissions for people with diabetes

- KPI 7: Identification of improved outcomes if certain ‘data insight’ based policies are implemented
- KPI 8: Identification of an EcoSystem of policy based services for improved patient outcomes and the identification of partners to drive improved outcomes

2 Pilot Requirements

This section describes the agreed Stakeholders for each pilot with Scenarios and User Stories which the agile UX testing will be based on. It also reports the new visualisation and UI requirements in addition to what was defined in D5.1. These were identified in collaboration with the specialists of each region who are part of the MIDAS Policy Board. Requirements reported here are the base for the UI development and testing. These requirements follow the KPIs defined in D5.1 for each MIDAS Pilot. The KPIs are taken into account with UX testing of MIDAS Platform.

Due to high variance between pilot topics, data and needed analytics, also the needs and requirements of Policy Boards varies from case to case. Some pilots had from the very beginning already a clear vision about their needs from visual and usability point of view when other cases have now the very first time a chance to see their data brought together and thus they were not yet able to define how they want to visualise or study the content of their data. As MIDAS develops single platform to be deployed on each pilot and in practice the differences between pilots are within used data and dedicated analytics, MIDAS can share the best practices to visualise the data over different pilots. We can share visualisation approach developed for a single pilot topic to other pilots, which can then bring fresh views for these pilots to study their own data.

2.1 Basque Country

Childhood obesity has emerged as an important public health problem in Europe and other countries in the world, and according to the world health organization (WHO), the worldwide prevalence of obesity nearly doubled between 1980 and 2008. In the WHO European Region 1 in 3 11-year-olds is overweight or obese (Health Behaviour In School-Aged Children (HBSC) Study)[1]. According to a childhood obesity review report[2], the increasing prevalence of childhood obesity is associated with emergence of comorbidities previously considered to be “adult” diseases (e.g. type 2 diabetes mellitus, hypertension, or nonalcoholic fatty liver disease), tracks strongly into adulthood, and it is the consequence of an interaction among a complex set of factors that are related to the environment, genetics, and ecological effects such as the family, community, and school. Basque Government, into its general lines of the XI legislature, within the health promotion and prevention action line pointed the creation of specific child obesity prevention plan as one of the new two main action areas

[1] http://www.euro.who.int/en/health-topics/Life-stages/child-and-adolescent-health/health-behaviour-in-school-aged-children-hbsc/growing-up-unequal.-hbsc-2016-study-20132014-survey
[2] http://www.mayoclinicproceedings.org/article/S0025-6196(16)30595-X/fulltext

(details in Basque[3] or Spanish[4]). As such, childhood obesity prevention has been targeted as prioritised regional policy for the Basque region within the MIDAS project.

2.1.1 Stakeholders

In the Basque Country, the identified stakeholders for the UX testing are: (i) Vice-Minister of Public Health, (ii) Province-level public health epidemiology managers and, (iii) Pediatrician responsible for an assigned children patient quota.

As the nomination of test users for each role is still underway, in the following part we present user personas that the test users will likely reflect. It is also probable that in the process of test user nomination, new test users might reflect additional persona types other than the ones mentioned below.

Name	Josune	**BIO**
Age	54	Education: Business Studies. Master in Business Administration (MBA). Background experience in managing Public Health teams. She also has past experience in managing private company. She has a quite tight agenda of government meetings, health commissions and coordination meetings with province level public health epidemiologists. She has to manage many public health policies, so time and involvement on each topic is not what she would like to. She needs synthesised information to understand each prioritised health issues' status, intervention areas and policy status. She also need easy to convey synthesised reports, to build trust with society and other political parties.
Occupation	Vice-Minister of Public Health	
Status	Married	
Location	Vitoria, Spain	
Occupation category		
primary expertise		
Quote "If I had simple and visual reports to explain up-to-date data-grounded descriptives, main candidate intervention areas, forecasted and actual implemented policies effect, government could take better decisions and reason them".		

3 http://www.euskadi.eus/eusko-jaurlaritza/-/albistea/2017/pertsonena-osakidetzako-langileena-eta-euskal-osasun-sistema-publikoa-unibertsal-zuzen-eta-kalitatezko-gisa-sendotzearena-izaten-jarraituko-du-xi-legealdiak/

4 http://www.euskadi.eus/gobierno-vasco/-/noticia/2017/la-xi-legislatura-seguira-siendo-la-de-las-personas-los-profesionales-de-osakidetza-y-la-de-la-consolidacion-del-sistema-publico-vasco-de-salud-como-universal-equitativo-y-de-calidad

MOTIVATIONS		PERSONALITY (%)	Choice with %
Incentive		Extrovert – Introvert	Extrovert 70%
Fear		Sensing – Intuition	Sensing 75%
Achievement		Thinking – Feeling	Thinking 70% - feeling 30%
Growth	Yes (minor motivation)	Judging – Perceiving	Judging 75%
Power		**TECHNOLOGY**	
Social	Yes (major motivation)	Average computer usage per week	25 hours
		Average internet usage per week	15 hours
GOALS		Average level of computer literacy (1-7)	4
a. To take data-grounded decisions, but having main intervention areas already synthesised		Familiarity with data analytics tool (e.g. SPSS, Excel, R, Matlab) (1-7)	2
b. To reason the adopted data-grounded policy decision		Familiarity of visualizations usage for decision making (1-7)	4
c. To present data-grounded policy forecast			
d. To present evolution of implemented policies		**Data analytics skills**	
FRUSTRATIONS		Integration(1-7)	3
a. Not up-to-date and population level data.		Analytics (1-7)	3
b. Lack trust of in adopted policies		Visualization (1-7)	4

c. Lack of reasoning of adopted policies	Interpretation (1-7)	6
d. Lack of simple and visual means to convey data in health commission and to society in general		

<table>
<tr><td>Name</td><td>Mikel</td><td colspan="2">BIO</td></tr>
<tr><td>Age</td><td>46</td><td colspan="2" rowspan="7">Education: Master degree in Epidemiology, Bachelor degree on Mathematics. Background experience in managing cancer and mortality registry, responsible for specific health issue managing within public health. Given the many possible areas of intervention of public health, actually he needs to find, get access, understand and cross different data sets before he starts interpreting the available data. Many of the data that he can access has been captured in limited surveys. He is willing to work with tools that would point him out candidate intervention areas and variables, and will allow him to forecast policy options and monitor implemented policy's evolution.</td></tr>
<tr><td>Occupation</td><td>Province-level public health epidemiology manager</td></tr>
<tr><td>Status</td><td>Single</td></tr>
<tr><td>Location</td><td>Bilbao, Spain</td></tr>
<tr><td>Occupation category</td><td></td></tr>
<tr><td>primary expertise</td><td></td></tr>
<tr><td colspan="2">Quote
"If the technology would point me out candidate intervention areas and patterns within the massive health, environmental, individual-generates, social,... data,”</td></tr>
<tr><td colspan="2">MOTIVATIONS</td><td>PERSONALITY (%)</td><td>Choice with %</td></tr>
<tr><td>Incentive</td><td></td><td>Extrovert – Introvert</td><td>Extrovert 60%</td></tr>
<tr><td>Fear</td><td></td><td>Sensing – Intuition</td><td>Sensing 85%</td></tr>
<tr><td>Achievement</td><td></td><td>Thinking – Feeling</td><td>Thinking 80% - feeling 50%</td></tr>
</table>

Growth	Yes (minor motivation)	Judging – Perceiving	Judging 75%
Power		**TECHNOLOGY**	
Social	Yes (major motivation)	Average computer usage per week	35 hours
		Average internet usage per week	20 hours
GOALS		Average level of computer literacy (1-7)	6
a. To have rate-adjusted detailed descriptives of health issues		Familiarity with data analytics tool (e.g. SPSS, Excel, R, Matlab) (1-7)	3
b. To cluster population in intervention profiles		Familiarity of visualizations usage for decision making (1-7)	2
c. To identify intervention areas and variables			
d. To forecast and evaluate policy outcome			
e. To monitor implemented policies evolution		**Data analytics skills**	
FRUSTRATIONS		Integration(1-7)	2
a. Working with data limited to samples/ survey-based		Analytics (1-7)	4
b. Difficulties to analyse heterogeneous data variables		Visualization (1-7)	5
c. Lack of support identifying candidate intervention areas / identifying new trends / patterns		Interpretation (1-7)	6
d. Lack of tools to forecast/evaluate policy options			

<table>
<tr><td>Name</td><td>Lourdes</td><td colspan="2">BIO</td></tr>
<tr><td>Age</td><td>38</td><td colspan="2" rowspan="7">Education: Medical degree and a specialisation on podiatry. 10 years pediatrician consultant experience and has participated in health research projects. She is concerned of classical activity and food recommendations limited effect and the increase of overweight and obesity cases. Even though she doesn't know her group compares to other nearby groups or general Basque Country indicators. She's willing to actively contribute to intervention, but it's not clear to her where to intervene besides classical recommendations of which other factors to capture in consultation to try to contribute to define issue interventions.</td></tr>
<tr><td>Occupation</td><td>Pediatrician responsible of an assigned children patient quote</td></tr>
<tr><td>Status</td><td>Married</td></tr>
<tr><td>Location</td><td>Bilbao, Spain</td></tr>
<tr><td>Occupation category</td><td></td></tr>
<tr><td>primary expertise</td><td></td></tr>
<tr><td colspan="2">Quote
"It's not clear to me if the overweight and obesity issue within my group is low or high in region / province or country context, and which are the intervention areas I could contribute"</td></tr>
<tr><td colspan="2">MOTIVATIONS</td><td>PERSONALITY (%)</td><td>Choice with %</td></tr>
<tr><td>Incentive</td><td></td><td>Extrovert – Introvert</td><td>Extrovert 80%</td></tr>
<tr><td>Fear</td><td></td><td>Sensing – Intuition</td><td>Sensing 75%</td></tr>
<tr><td>Achievement</td><td></td><td>Thinking – Feeling</td><td>Thinking 70% - feeling 40%</td></tr>
<tr><td>Growth</td><td>Yes (minor motivation)</td><td>Judging – Perceiving</td><td>Judging 55%</td></tr>
<tr><td>Power</td><td></td><td>TECHNOLOGY</td><td></td></tr>
<tr><td>Social</td><td>Yes (major motivation)</td><td>Average computer usage per week</td><td>20 hours</td></tr>
</table>

		Average internet usage per week	15 hours
GOALS		Average level of computer literacy (1-7)	5
a. To know her group overweight/incidence within region/province/country		Familiarity with data analytics tool (e.g. SPSS, Excel, R, Matlab) (1-7)	3
b. To know which other variables she can capture for further analysis		Familiarity of visualizations usage for decision making (1-7)	4
c. To know which specific intervention areas to work in her group/context		**Data analytics skills**	
FRUSTRATIONS		Integration(1-7)	2
a. Lack of updated picture of her group state within at basque context.		Analytics (1-7)	4
b. Lack of group specific intervention areas		Visualization (1-7)	4
		Interpretation (1-7)	5

2.1.2 Scenarios and User Stories

The actual scenarios and user stories are being researched and developed within MIDAS in the context of a specific Health Plan which is in development to tackle child obesity. There are also other related government initiatives as “Healthy Nutrition Plan” recently introduced. The Health Promotion, Health Surveillance, Health Protection policy processes are involved in the following scenarios.

2.1.2.1 Descriptive or Effectiveness assessment scenario

Nowadays, the reality of children overweight and obesity is captured through surveys covering a limited sample and a limited time granularity. Health managers lack of up-to-date detailed descriptives of the population. Moreover, they cannot do a reliable analysis of policy implementation pilots effectiveness.

In the Basque autonomous country, the most comprehensive obesity data about children's obesity available nowadays comes from the Health Survey. This survey is made every 5 years and it consists in a large questionnaire about many health topics, completed by a sample of the population. The results of the survey are presented as written reports, where you can find static table and graphs. The last reports where published in 2015 and the came from the survey of 2013. Other specific studies are also done but are limited in specific areas and / or ages.

User story

The main user story within descriptive or effectiveness assessment scenario, is that of health managers at different levels (Individual => Patients-group assigned Pediatrician (Patient quote) => Province-level public health epidemiologist => Country-level public health manager) that can have an updated descriptive of their aggregation level and a contrast with others at their aggregation level or at higher levels of aggregation. Additionally, visualising the evolution of these descriptive visualizations with implemented pilot policies would help, for further considering them.

2.1.2.2 (Inter) variable correlation or relationship identification scenario

Currently, it's not clear which factors are responsible or affect in the process of a children evolving to a overweight or obese state. Therefore, intervention areas are not clear and interventions are non-data (or limited-data) grounded. The interventions usually target intervention areas that are established within general practice, reported in the scientific state of art, or hypothesised by experts' experience and knowledge.

User story

Identification of relationships (one to one and more complex ones) among predictor variables (i.e. health and non-health variables) and variables representing the children overweight and obesity. Visualisation of identified relationship (e.g. through correlation analysis, or regression) to domain experts, so they can concentrate on some interventions areas, and choose those that make sense (avoiding non-spurious correlations) and are targetable by the Public Health manager. An example visualisation can be a correlation matrix. Additionally, it would be desirable to identify clusters of children that have similar profile and patterns that lead to similar values of variables representing children overweight and obesity.

2.1.2.3 Forecast or Monitor policy scenario

An evolution of the available tools to the properties introduced for the descriptive user story, will allow to respond to "**What's the current status?**" question. The next evolution of the tools to the properties presented for the variable correlation / relationship identification user casestory, will allow to respond to "**Which should be my policy's intervention areas?**". But these do not allow for predicting the evolution of children's overweight and obesity health issue, nor analysing data model-based policy resulting scenarios

User story

This user story will allow health managers to forecast short-medium-long term evolution of children's overweight and obesity health issue, based on historical data and data being updated. This user story will also target data model-based forecasting of scenarios, where intervention policies can be chosen based on policy-expected predictor variables' value improvements, which lead to optimum future scenarios. Additionally, monitoring of the evolution of an implemented data model-based policy could be targeted for feedback.

2.1.3 Visual Requirements

Currently, stakeholders in the Basque Country use some simple and static graphs for analyzing the situation and to make decisions.

- Public Health Manager / Director of Public Health and Addictions

In order to present and report the public health situation in health commissions and to the public in general, Public Health Manager / Director of Public Health and Addictions uses Microsoft Office Powerpoint presentations displaying public health status by means of line and bar charts.

- Province-level Public Health Epidemiologist

With the aim of analyzing the origin of health issues and to act on them, Public Health Epidemiologist uses an Excel-based dashboard. These visualisations are mainly vertical histograms for incidence rate among gender per each age-range and tables showing the number of cases by zones and age ranges. In the table, the values that have unusual evolution are coloured in red. All this data is rate-adjusted by age. This dashboard is used in other health issues, but it is not available for children obesity yet.

- Patient-group assigned pediatrician

Pediatrician uses individual growth charts for tracking the physical evolution of his patients. Growth charts consist of a series of percentile curves that illustrate the distribution of selected body measurements (height and weight) in children.

These visualizations are quite limited. No difference has been found for analytics and visualization for people with different social status nor visualizations that allow the comparison of BMI with other variables.

The proposed visualizations are the following (per user story):

<u>Descriptive visualisations User Story</u>

- Map-based visualisation with different aggregation level (Provinces, Regions vs Trust / OSI, municipalities, assigned patient-quota area) visualisation, e.g. Choropleth
- Represent differences among regions.

- Represent positive / negative evolution in time.
- Rate-adjusted visualization
- Analytic applied to variable before being plotted.
- Line-chart with multiple lines, mark average and see which below, which above for different aggregation levels.
- Percentile graph for different aggregation levels.
- Scatterplot linked to map visualisation.

Variable correlation / relationship identification user story

- Correlation matrix
- Cluster visualisation
- Scatter-plot with PCA
- Radar-chart

Forecast / monitoring User story

- Short-medium-long forecast of variables representing child overweight and obesity
    - Use of the same visualisations as from descriptive and effectiveness assessment, but overlaying, following the line in a different colour for forecasts.
- Exploration of data model-based alternate scenarios, based on predictor variables options testing
    - Editor of sets of predictor variable configurations, that get plotted as short-medium-long forecast of variables' targeted visualisations
- Monitoring of implemented policies
    - Alerts within dashboard if current (or updated short-term prediction) child overweight and obesity deviations occur from the implemented policy forecast.
    - Visualisation through overlays and / or in lines with different colours of implemented policy forecasted, what happened until now and the new forecast, based on updated data

2.2 Finland

2.2.1 Stakeholders

"Preventive mental health and substance abuse of young people" as the Finnish pilot topic was established based on multiple interviews and workshops with policy makers from the city level, regional level and national level in Finland. From the early stage of the MIDAS project, the Finnish consortium has maintained continual communication with stakeholders from various levels in order to understand general policy making user requirements when it comes to utilization of rich data. For the next phase of the MIDAS project, it has been agreed in the agile iterative schedule for policy board engagement and platform testing, we are in the process of nominating test users from different levels.

While nominating test users for the iterative training and usability testing rounds, we are considering to engage policy board members from different roles as well. These roles include: 1) City/Regional level policy maker, 2) Data scientist, 3) Clinical professional/ Researcher, 4) National level policy maker. As the nomination of test users for each role is still underway, in the following part we present user personas that the test users will likely reflect. It is also probable that in the process test user nomination, new test users might reflect additional persona type than the ones mentioned below.

City level policy maker persona

<table>
<tr><td>Name</td><td>Jenni</td><td colspan="2">BIO</td></tr>
<tr><td>Age</td><td>55</td><td colspan="2" rowspan="7">Jenni is the director of health services, with many years of experience in the field. She has a background as a doctor with basic training. Currently, she is responsible for organizing health services, mental health & substance abuse services in city level. Additionally, she participates as an expert in the national healthcare reform project from a regional level also. In personal life, Jenni has twin daughters (15 year olds) and an elder son (20 year old). She contemplates various necessities for mental health case comparing in her own experience and her kids teenage growing up.</td></tr>
<tr><td>Occupation</td><td>Director of health services, City of Oulu</td></tr>
<tr><td>Status</td><td>Married</td></tr>
<tr><td>Location</td><td>Oulu, Finland</td></tr>
<tr><td>Occupation category</td><td>Policy</td></tr>
<tr><td>primary expertise</td><td></td></tr>
<tr><td colspan="2">Quote
We don’t have information about cost and impact yet and what is the time window during which we can see some investment, it’s always said that one euro invested in substance abuse services saves six euros but what is the time window there.</td></tr>
<tr><td colspan="2">MOTIVATIONS</td><td>PERSONALITY (%)</td><td>Choice with %</td></tr>
<tr><td>Incentive</td><td></td><td>Extrovert – Introvert</td><td>Extrovert 80%</td></tr>
<tr><td>Fear</td><td></td><td>Sensing – Intuition</td><td>Sensing 75%</td></tr>
</table>

Achievement		Thinking – Feeling	Thinking 70% - feeling 30%
Growth	Yes (minor motivation)	Judging – Perceiving	Judging 80%
Power		**TECHNOLOGY**	
Social	Yes (major motivation)	Average computer usage per week	20 hours
		Average internet usage per week	10 hours
GOALS		Average level of computer literacy (1-7)	4
a. Organizing health services in city level (resource allocation)		Familiarity with data analytics tool (e.g. SPSS, Excel, R, Matlab) (1-7)	2
b. Organizing mental health & substance abuse services in city level		Familiarity of visualizations usage for decision making (1-7)	4
c. Facilitate the national healthcare reform from regional level		**Data analytics skills**	
FRUSTRATIONS		Integration(1-7)	3
a. Most of the relevant data is in silos, integration from multiple sources is a challenge.		Analytics (1-7)	2
b. generalized preventive mental health services might hinder target group specification		Visualization (1-7)	3
c. Lack of impact measurement of services with existing tools		Interpretation (1-7)	6

Regional level policy maker persona

Name	Pia Majala	**BIO**
Age	58	Ms. Majala is a regional politician who is currently working as the Director of Education for Kuusamo region, Finland. She has been involved in public service for more than 30 years. Her work mainly focuses on organizing education related services, but at the same time she contributes to the physical and mental well-being
Occupation	Municipal Politician, Director of Education	
Status	Married	
Location	Kuusamo, Finland	

Occupation category	Policy	of young people as a politician. She has academic degree specialising in Education. Also, she has worked in multiple projects focusing on wellbeing and mental health in her long public service career.	
primary expertise			
Quote I feel like utilising data, is still today, not systematic in municipal decision-making. It's pretty separate, so when I make preparations there are certain things of course, I utilise it for certain decision making situations. But, it's not systematic in any way.			
MOTIVATIONS		**PERSONALITY (%)**	Choice with %
Incentive	minor	Extrovert – Introvert	Extrovert 70%
Fear		Sensing – Intuition	Intuition 40%
Achievement	minor	Thinking – Feeling	Thinking 50%
Growth		Judging – Perceiving	Perceiving 30%
Power	Major	**TECHNOLOGY**	
Social	Major	Average computer usage per week	20 hours
		Average internet usage per week	12 hours
GOAL		Average level of computer literacy (1-7)	4
a. Collect relevant data from multiple sources and make sense out them through comparative analysis.		Familiarity with data analytics tool (e.g. SPSS, Excel, R, Matlab) (1-7)	2
b. To create a more data oriented decision making practice in the regional level.		Familiarity of visualizations usage for decision making (1-7)	5
c. To identify and utilize more variables that are relevant to youth wellbeing, besides education data and financial figures.		**Data analytics skills**	
FRUSTRATIONS		Integration (1-7)	4
a. Lack of comprehensive dataset and up to date information on youth people.		Analytics (1-7)	2

b. Currently decision making is perhaps based more on “present moment”, not always evidence oriented and futuristic.	Visualization (1-7)	3
c. Resources and tools at municipal level is not substantial enough.	Interpretation (1-7)	5
d. Most of the decisions are not made based on health/ wellness issues, rather focuses more on economic factor.		

Data scientist/ researcher persona

Name	Maija	**BIO**	
Age	35	Maija is a researcher working at THL. She has spent all of her work career at THL and even did her practical training there, before graduation. Her work mainly focuses on scientific research on social group differences in health determinants such as education, income and health behaviour. She has academic degree specialising in Health Sciences. She is married and has two children.	
Occupation	Researcher, THL		
Status	Married		
Location	Helsinki, Finland		
Occupation category	Technical, data specialist		
primary expertise			
Quote Information on health and welfare goes beyond static figures and graphs. It is important to be able to examine self defined sets of variables, and make estimates, based on reliable, evidence based data.			
MOTIVATIONS		**PERSONALITY (%)**	Choice with %
Incentive		Extrovert – Introvert	Introvert 70%
Fear		Sensing – Intuition	Sensing 80%
Achievement	Yes	Thinking – Feeling	Thinking 60%
Growth	Yes	Judging – Perceiving	Perceiving 60%
Power		**TECHNOLOGY (level)**	
Social	Yes	Average computer usage per week	30 hours
		Average internet usage per week	10-15 hours
GOALS		Average level of computer literacy (1-7)	6

a. Spotting differences between social groups and addressing directives at them.	Familiarity with data analytics tool (e.g. SPSS, Excel, R, Matlab) (1-7)	7
b. Trying to make the directives be included in policies.	Familiarity of visualizations usage for decision making (1-7)	5
c. Trying to make her message noticed on higher levels of administration.	**Data analytics skills**	
FRUSTRATIONS	Integration(1-7)	5
a. Cooperation even within her own institute is challenging.	Analytics (1-7)	4
b. Processes are slow and somewhat stagnated.	Visualization (1-7)	5
c. Means of communication often insufficient.	Interpretation (1-7)	7

Clinical professional / researcher persona

Name	Anna-Mari	**BIO**	
Age	38	MD, PhD, Adjunct Professor Anna-Mari works as a researcher in University of Oulu and psychiatrist in Oulu University Hospital. She has studied schizophrenia for nearly 2 decades, and she is specially interested in epidemiology, outcomes, cognition, and treatment in psychoses and other mental disorders. Currently she focus on research related to effects of antipsychotics.	
Occupation	physician, researcher		
Marital Status	Married		
Location	Oulu, Finland		
Occupation category	Research		
primary expertise	Epidemiology, research		
Quote			
MOTIVATIONS		**PERSONALITY (%)**	Choice with %
Incentive		Extrovert – Introvert	both, more extrovert about 70%

Fear		Sensing – Intuition	both, more sensing about 70%
Achievement	some	Thinking – Feeling	both, more thinking about 60%
Growth	some	Judging – Perceiving	?
Power		**TECHNOLOGY (level)**	
Social	some	Average computer usage per week	20 hours
		Average internet usage per week	10 hours
GOALS: Personal / occupational goals in the context of health policies.		Average level of computer literacy (1-7)	7
a. To understand causes and consequences of mental disorders, to participate in enhancing the treatments and rehabilitation for mental disorders (research)		Familiarity with data analytics tool (e.g. SPSS, Excel, R, Matlab) (1-7)	6
b. To effect the development of mental health services in long-term (via implementation of research)		Familiarity of visualizations usage for decision making	1
		Data analytics skills	
FRUSTRATIONS: Personal/ occupational frustrations in the context of health policies.		Integration (1-7)	5
a. slowness of implementation of research/ evidence base to health care services		Analytics (1-7)	6
b. poor availability of mental health care services locally		Visualization (1-7)	5
		Interpretation (1-7)	6

Medical researcher persona

Name	Heini Korhonen	**BIO**
Age	56	

Occupation	Adjunct professor, University lecturer	Heini Korhonen works as an Adjunct Professor and University Lecturer in the University of Oulu, Finland. Her main research areas are neuropsychiatric disorders in children, adolescents, and adults. She leads several longitudinal epidemiological studies in Northern Finland. She is married and has three adult children.	
Marital Status	Married		
Location	Oulu, Finland		
Occupation category	Research		
primary expertise	Medical Research, longitudinal epidemiological studies		
Quote			
MOTIVATIONS		**PERSONALITY (%)**	Choice with %
Incentive		Extrovert – Introvert	Extrovert 40%, Introvert 60%
Fear		Sensing – Intuition	Sensing 50%, Intuition 50%
Achievement	Yes	Thinking – Feeling	Thinking 40%, Feeling 60%
Growth		Judging – Perceiving	Judging 30%, Perceiving 70%
Power		**TECHNOLOGY (level)**	
Social	Yes	Average computer usage per week	30 hours
		Average internet usage per week	20 hours
GOALS: Personal / occupational goals in the context of health policies.		Average level of computer literacy (1-7)	6
a. Implementation of basic research into health policy		Familiarity with data analytics tool (e.g. SPSS, Excel, R, Matlab) (1-7)	5
		Familiarity of visualizations usage for decision making	2

	Data analytics skills	
FRUSTRATIONS: Personal / occupational frustrations in the context of health policies.	Integration (1-7)	5
	Analytics (1-7)	5
	Visualization (1-7)	3
	Interpretation (1-7)	6

National level policy facilitator (development manager)

Name	Tapio	**BIO**	
Age	38	Tapio is a development manager of secondary use of data. His duties and interests include strategic data resources development including open data and other data products, data policy implementation, development of data services and customership, reporting and presentation of data, and UX. He also has a lot of experience of international projects in developing public health monitoring by data.	
Occupation	Development manager, National Institute for Health and Welfare (THL)		
Marital Status			
Location	Helsinki, Finland		
Occupation category	Strategic development (policy)		
primary expertise	Secondary use of data (open data, data analytics, data visualisations)		
Quote	No matter how important your work is as such – if you can't make it seen, it's all futile.		
MOTIVATIONS		**PERSONALITY (%)**	Choice with %
Incentive		Extrovert – Introvert	70 - 30
Fear		Sensing – Intuition	70 - 30

Achievement	Yes	Thinking – Feeling	80 - 20
Growth	Yes	Judging – Perceiving	10 - 90
Power		**TECHNOLOGY (level)**	
Social		Average computer usage per week	35 hours
		Average internet usage per week	10 hours
GOALS: Personal / occupational goals in the context of health policies.		Average level of computer literacy (1-7)	6
a. Right data to be chosen to describe real phenomena.		Familiarity with data analytics tool (e.g. SPSS, Excel, R, Matlab) (1-7)	4
b. Accurate and timely data to be available.		Familiarity of visualizations usage for decision making	6
		Data analytics skills	
FRUSTRATIONS: Personal / occupational frustrations in the context of health policies.		Integration (1-7)	3
a. Silos, even within own organisation.		Analytics (1-7)	4
b. Naysayers.		Visualization (1-7)	5
C. Lack of expertise.		Interpretation (1-7)	7

National level policy facilitator (Research professor)

Name	Pekka Juga	**BIO**
Age	57	Pekka Juga holds professorships and faculty appointments at the National Institute for Health and Welfare, and multiple universities in Finland and Sweden. His main research focus has been in utilization of routinely collected health and welfare registers. These include the Medical Birth Register, Hospital Discharge and Outpatient Register,
Occupation	Research professor	
Marital Status	Cohabiting	

Location	Helsinki, Finland	Prescription Register and Cause-of-Death Registers with experience using registers in all Nordic countries and in several European countries. The main research interests include perinatal and maternal health, migrant health, childhood and adolescent health, sexual and reproductive health, mental health, and use of health care services. Increasingly, the studies have included a longitudinal component, usually from the prenatal period until adulthood.	
Occupation category	Research and statistics (policy)		
primary expertise	Register-based research		
Quote	In demography, a quarterly is 25 years.		
MOTIVATIONS		**PERSONALITY (%)**	Choice with %
Incentive		Extrovert – Introvert	70 - 30
Fear		Sensing – Intuition	65 - 35
Achievement	Some	Thinking – Feeling	70 - 30
Growth	Some	Judging – Perceiving	25 - 75
Power		**TECHNOLOGY (level)**	
Social	Some	Average computer usage per week	40 hours
		Average internet usage per week	40 hours
GOALS: Personal / occupational goals in the context of health policies.		Average level of computer literacy (1-7)	6
a. Increase the use of Finnish register data in scientific research		Familiarity with data analytics tool (e.g. SPSS, Excel, R, Matlab) (1-7)	6
		Familiarity of visualizations usage for decision making	5
		Data analytics skills	
FRUSTRATIONS: Personal / occupational frustrations in the context of health policies.		Integration (1-7)	6
a. Low use of information		Analytics (1-7)	7

b. Excuses to hinder register-based research	Visualization (1-7)	5
	Interpretation (1-7)	6

2.2.2 Scenarios and User Stories

In deliverable 5.1 we briefly presented the scenarios and user stories arising from the Finnish MIDAS pilot, which are titled as regional scenario and national scenario. We focus on the same scenarios in the next phase of the project, however in this deliverable we tend to describe scenario with a focus which assists in defining different tasks for the formative iterative usability tests for the MIDAS platform. At first, we describe the scenarios from a general level and in the user stories we raise different perspectives that users of MIDAS platform may face in their jobs.

The chapters below also comprehend the different level of users of the MIDAS platform. So, the user stories will combine both perspectives of data scientists and also policy makers.

2.2.2.1 Regional scenario:

In the regional case, the Finnish pilot focuses on more short to medium term policy implication scenarios, which are in most cases handled by regional administrative or policy bodies. Although the ministry of social affairs and health is most concerned about health policies, since Health is such a subject that crosses multiple boundaries of contemporary organizations, often different other administrative bodies in the regional level is observed to be carrying out policy programmes which affects public health in one way or the other. More importantly, this is more applicable when the Finnish pilot is focusing on preventive mental health of young people and intoxicant abuse problems. During the earlier interviews and workshops conducted for the Finnish pilot, it reflects that policy activities from health and healthcare services, social services, education services, social welfare services, and even the city office have direct or indirect roles to play in deploying preventive programmes for mental health of young people.

The key research question that we address in this scenario is:
How to use MIDAS platform and rich data from various sources in different regional policy organizations to support preventive policy making?

The research question shows the inclusion of different regional policy organizations in order to coherently implement high level policies, since the pilot topic of preventive mental health is a broad one. The regional scenario includes stakeholders from city and regional levels; meaning: the stakeholders from City of Oulu, Kuusamo, Kempele, Oulunkaari and their decision making situations are more directly considered in this scenario. As the regional scenario focuses on supporting preventive policy making, currently these decisions are made based on national indicators, periodic information sourced from partners/ paid vendors (e.g. monthly excel based reports from addiction care services), primary data collected from meetings and customer orientation (mostly qualitative), and OukaDW for some specific cases.

Some of the key decisions which are highlighted are: financial resource allocation for preventive programmes, personnel resource allocation, new campaign design and launch, etc. The Social care and Health and welfare services of the City of Oulu also generates high level follow up reports periodically (every 6 months) to support decision making. The challenge is, regional decision makers face the obstacle arising from confidentiality and data protection. While the amount of data is yet not so significant in most of the stakeholder site, individuals are possible to be identified which creates the conflict against confidentiality and data protection. So, the regional case in Finnish MIDAS pilot will aggregate and process data from multiple sources by removing all identifiable variables and securing all highly sensitive variables.

User stories:

In the regional scenario, we identify that multiple user types will be using the MIDAS platform in practice. They can be regional level policy maker, data scientist from the regional policy organization, clinical professional/ researcher working on a regional case, etc.

User stories for each of the stakeholders in the regional scenario will differ to some extent. However, depending on the data sources they opt to use for analytics and visualizations, this might be limited to common analytics and visualizations to describe the possible estimations in the short term future e.g. trends for future. If the regional level policy maker and data scientist at policy organization personas are considered to work together, then data scientists will do much of the data analytics and create correlations and heatmaps. Also, the data scientist will append some comments and remarks to the dashboard which will make it more useful for the policy maker's clarification.

As the Finnish Pilot is including data from Northern Finland Birth Cohort 1986, THL data repository and City of Oulu's data, stakeholders can aim to identify correlations among different variables on mental health in general and wellbeing. For example, the Northern Finland Birth Cohort dataset includes six different survey results which have substantial number of respondents for every surveys. These surveys also involve numerous useful variables to create analytics with. However, one challenge for users would be to thoroughly be informed about the variables of the dataset before they perform analytics. Practically, this becomes an issue for high level policy makers who might have limited time to devote themselves for these activities.

Data source and variable types are also very important when it comes to clinical professional and researcher persona. In practice, clinical professionals or researcher will be able take benefit from the MIDAS platform in the case if the data in the platform is related to their topic of study. However, since the datasets that are being brought in to Finnish pilot are quite extensive, proper examination of the data sets will enable the clinical professionals to identify relevant databases to their studies.

2.2.2.2 National scenario:

As mentioned, the Finnish pilot topic being quite broad makes the need to look beyond the boundaries of conventional health and wellbeing policy organizations. Although such organizations are in the centre of preventive policy making, for implementation other policy wings need to be involved. The focus of the national case is long-term and strategic. By introducing data-driven foresight to the national level policy players, Finnish MIDAS pilot intends to justify the value of rich data in long-term policy making.

The main research question for the national scenario is:
What kind of regulations, enablers and metrics should be enforced and how in order to achieve the systemic targets escalating on various government sectors?

Using system dynamics modeling, a draft model for the preventive mental health for young people and intoxicant abuse has been created through multiple workshops with national level policy makers through experiential knowledge. Once validated, this model can be used to simulate different kinds of variable changes in social and personal level, including variables which can be impacted by policy implementation. In the national scenario of Finnish MIDAS pilot, the data from different sources can be used to identify correlations between different variables based on actual population data.

Currently, in Finland, the healthcare sector is in the middle of a huge change due to the social and healthcare reform. It is going to change the responsibility of organizing services and financial resources of healthcare from 326 municipalities and 20 hospital districts to 18 counties. Currently each municipality and hospital district have their own decision support systems and the way these systems are built varies a lot. Also, there is an ongoing development to modify legislation related to the use of different data sources on a national level. Based on the interviews, we understand that the new legislation is constructed in a way, which is compatible to EU data protection regulation (GDPR).

The availability of data is not the challenge on the national level, but refining it for proper application is. Since the early 90's in Finland, municipalities have been responsible for all of it, including the most demanding specialist health care. According to the Finnish constitution, municipalities have the general authority on things that they are legally required to provide. This means that on the state level, there has been no need and in fact not even a mandate, for more detailed data-based governing.

User stories:

Similar to the regional scenario, there will be multiple types of users of MIDAS platform on the national scenario as well. They can include: National level policy maker, data scientist in ministry or similar organizations, data scientist/ researcher in THL or similar statistics and data repository organizations.

For the preventive mental health case, it is assumed that policy activities can take place simultaneously in different ministry level organizations. In that essence, communication among these organizations will be vital to manage congruence and achieve cohesive preventive policies.

Also, on a national level, decision-making involves individuals involved with political parties and apolitical civil servants supporting them. Decision-making foregrounds differ sometimes within individuals with varying professional goals. Decision-making strategies can also vary for a politician over time, depending on personal and political goals in respect to other parties. Typically, materials developed and brought by apolitical civil servants support legislative decisions. In this context, strategic foresight through system dynamics modeling and simulation will prove to be a handy tool to create legislation that does not only protect political goals, but also advances preventive decision making in regional levels. In addition to strategic foresight, it has been acknowledged that more suitable analytics and visualization techniques can foster policy formulation at the national level, creating a coherent overall scenario by connecting it to the regional cases.

Returning to the Finnish pilot and the MIDAS pilot, thus far, a system dynamic model (SDM) has been drafted for the preventive mental health of young people and intoxicant abuse case in cooperation with national level policy players. Though the model is currently under validation, during the UX test sessions causal relationships identified in the SDM can be tested through different analytics and visualization methods. As an example: the draft SDM depicts School performance of young people is positively related with the education level of their parents. However, test users can attempt to identify if there is any relationship of these variables with the mental illness prevalence statistics or psychological wellbeing in general. In the NFBC 1986 dataset, there are numerous variables related to hobby activities and leisure time of 15-16 year olds. This can be also tested to correlate how these hobby activities are affecting psychological well being.

2.2.3 Visual Requirements

Currently, visualization techniques are not widely used in stated stakeholder cases. While in some instances plot diagrams, pie charts, and population profile visualizations are being used, they are based on a limited amount of static data. At the moment, the visualizations that are applied are made with regularly available software packages like SPSS or MS Office suite. In some day-to-day cases, handmade scribbles and figures are often used as visualizations to convey data/message/knowledge from data collectors to upper layer decision makers. However, on the national level, a good set of visualization examples are provided by THL in SotkaNet[5] and other statistical services like TEAviisari[6], WelfareCompass[7] and Terveytemme[8], and similar would be required for MIDAS as well.

With regards to the MIDAS platform, from the initial user testing round, a key insight has been identified in the Finnish Pilot. It is assumed that at the top level of policy making, there are experienced policy makers who have been working in the sector for long periods. However, this can be too time consuming for these experienced policy makers apt to utilizing rich data first hand, and straightforward visualizations might not answer clearly to their enquiries. For this reason, the data scientist/ clinical professional/ researcher persona seems to be a proper fit to analyze and visualize different data sets and convey the results through the MIDAS platform through the "Share dashboard" function. Test users have proposed two vital improvement areas in user interface which can assist in such communication between top level policy makers and data scientist personas. They are:

1. When analytics are performed and visualizations are created, it would be useful to have some short text summaries explaining the highlights of the specific visualization.
2. In case of a data scientist creating a dashboard using different kind of analytics and visualization on behalf of a policy maker, they will share this dashboard to the policy maker. However, if the dashboard is shared without any meaningful comments and remarks, it does not serve the purpose to full extent. Currently, the communication channel between different users are limited, which can be improved by adding comment or discussion option within the dashboard. For

5 https://sotkanet.fi/sotkanet/en/index?

6 https://teaviisari.fi/teaviisari/en/index?

7 https://www.hyvinvointikompassi.fi/en/web/hyvinvointikompassi/

8 http://www.terveytemme.fi/

the Finnish pilot case, it will also be useful considering the necessary communication among different policy organizations, e.g. health, education, welfare, unemployment etc.

Based on the identified user stories and scenarios the following advanced visualization techniques are identified to be useful in the Finnish Pilot; either in national / regional scenario or in both scenarios, depending on the data on which the technique is applied. However, this list of visualization techniques is not exhaustive, rather it will evolve over time as the Finnish Pilot gets access to the data and variables requested from multiple sources:

● **Network diagram:** Cause and consequence analysis is identified as a key element, as it should help policy makers if better visualized. Network diagram is a visualization technique that shows relationship between numerous variables. It can display the causal relationships, perhaps indicate correlation, most importantly it shows the the ripple effects of decisions being made.

● **Area graphs:** Area graph is a useful visualization tool to display the development of quantitative values over an interval or time period. They are most commonly used to show trends. This type of visualization can be applicable for both the regional and the national scenario to reflect on longitudinal indicator data.

● **Stacked area graphs:** Stacked area graph is a modified version of the area graph where longitudinal data on multiple variables are presented in the same visualization. In this way, trends of multiple variables can be compared and possible correlations can be pointed out.

● **Pie chart:** Pie chart and donut charts are useful tools when making decision on resource allocation. Therefore, this visualization technique relates more to the regional scenario in the Finnish pilot.

● **Choropleth Map:** Choropleth Maps display divided geographical areas or regions that are coloured, shaded or patterned in relation to a data variable. This provides a way to visualise values over a geographical area, which can show variation or patterns across the displayed location. In the Finnish pilot, this technique can be used in both regional and national level. On the regional level, it could be used to portray area profiles to make it visually clear for regional decision makers for which locations in the region requires what type of attention. The regional level shall need to include the sub-regions of the city of Oulu to illustrate OUKA data. Similarly, on the national level, an overall national comparison can provide better understanding on specific issues. Note: here the aim to provide similar map functionalities as defined for Basque in chapter 2.3.1..

● **Timeline:** A Timeline is a graphical way of displaying a list of events in chronological order. The main function of Timelines is to communicate time-related information, either for analysis or to visually present a story or view of history. Especially in the Finnish national case, this technique can be used to reflect how different variables have changed over time and their effect in the present. Timelines can be used in parallel with Network diagrams to dig deeper into the causal relationships, where timelines portray in detail about separate variables. From another perspective, timelines are a very useful tool for the Preventive mental health to show how the situation has evolved over the time, which helps policy makers with decision making. Besides network diagrams, other graphs can be combined with a Timeline to show how quantitative data changes over time.

The above elements together with the common elements decided for the first iteration of the platform can also support the system dynamic simulations. Typically they require visualization of dynamic hypothesis diagrams, causal maps, time series curve sets, sensitivity graphs, stacked columns and bars. The user interface commonly also contains graphical tables for inserting what-if assumptions of different scenarios. The actual needs for system dynamic modelling forecasting tools visualizations will be defined based on the first iteration experiences during the second and third iteration.

2.3 Northern Ireland

2.3.1 Stakeholders

Role Title	**Role Description**
IT Leads (BSO)	The IT Leads will supply the Information Staff with the necessary data sources, data structure and refreshing of the data when required. They will help the Information Staff understand the data should any queries arise when it comes to creating the Dashboards on the MIDAS Platform.
Information Staff	This team is made up of professionally qualified Statisticians and will be responsible for creating dashboards on the MIDAS platform using the data provided by the IT Leads. They will create dashboards and build the content based on requests from the Policy Staff.
Policy Staff	Policy Staff will provide the Information Staff with requests for analysis on a specific dataset that an upcoming policy review will need. They will also interact with the Dashboard created by the Information Staff and they will explore the trends and patterns to help form their analysis.

Policy Lead	The Policy Lead will carry out presentations at Board Level to the Departmental Leads to provide evidence and underpin reasoning in Policy Making. The presentation will be based off of the discussions held with the Policy Staff regarding the information gathered from the MIDAS Dashboards and the analysis that was formed.
Departmental Leads	Departmental Leads will endorse or reject policy proposals based on the presentations and discussions with the Policy Leads.

PERSONA A: Policy Lead

NAME	Eilis
AGE	54
JOB ROLE	Director, Family and Children's Policy Directorate, Department of Health
LOCATION	Belfast
TRAITS	Results-focussed; people-oriented.

BIO	Education: BA in Celtic Languages and Literature; MSc. in Computer Science and a Masters in Business Administration. Background: Worked in Irish Place-name research in Queen's University Belfast at the start of her career. Following that, she worked as a software developer for around 10 years before moving into policy development in central government. Within the Department of Health, she has lead responsibility for policy and legislation relating to family support, child protection, looked after children and adoption. She is also leading on the implementation of an Early Intervention Transformation Programme – a cross-departmental initiative to provide early support to families and children in need. Previously, she worked in the Children and Young People's Unit in the former Office of the First and deputy First Ministers, where she was responsible for driving forward the overarching Northern Ireland 10-year Strategy for Children & Young People. She has also worked as Private Secretary to a former First Minister.
GOALS (objectives this person hopes to achieve through MIDAS)	The development of robust KPIs, improved measurement of outcomes; better designed services and provision; the right services in the right place; better informed policies and strategies to meet the needs of vulnerable children and young people, including those in care; benchmarking good practice across HSC services and HSC Trust areas and establishing consistency of approach; revised Delegated Statutory Functions reporting (i.e. more efficient monitoring and analysis).
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	DoH is restricted to 6 monthly and annual "snapshot" data (gathered at a single point in time) rather than longitudinal data. It is therefore difficult to establish patterns and historical trends with ease, making effective policy decision-making and responsive service planning and delivery more difficult to deliver.
TECHNOLOGY SKILLS (% level)	
IT and Internet	70%

Software	50%
Mobile Apps	50%
Social Networks	50%
DATA ANALYTICS SKILLS (% level)	
Integration	40%
Analytics	40%
Visualization	40%
Interpretation	40%

PERSONA B: Information Staff

NAME	Heidi
AGE	44
JOB ROLE	Deputy Principal Statistician
LOCATION	Department of Health, Belfast, Northern Ireland
TRAITS	Creative and supportive
BIO	Master in Psychology from the Norwegian University of Science and Technology Worked as a statistician for Department of Health the last 8 years, focussing on children's social care the last 5 years. Enjoys a 'hands-on' approach when working with datasets, to be able to see trends that cannot be picked up from running analysis or a report.

GOALS (objectives this person hopes to achieve through MIDAS)	Establish links between datasets that can set precedence for future work.
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	There is no fast-track, it always takes longer than expected to go from A to B, B to C, C to…
TECHNOLOGY SKILLS (% level)	
IT and Internet	70%
Software	70%
Mobile Apps	50%
Social Networks	50%
DATA ANALYTICS SKILLS (% level)	
Integration	70%
Analytics	80%
Visualization	80%
Interpretation	80%

PERSONA C: IT LEAD

NAME	Susan
AGE	55
JOB ROLE	Head of Regional Data Warehouse, HSCNI

LOCATION	Belfast
TRAITS	Optimistic; collaborative
BIO	Education – BSc in Physics and Computer Science. She has worked in Health Service IT for over 30 years. She has worked in application development and support for most of that time, with a particular interest in making information from operational systems available to users for reporting and analysis. She manages a small team within the Business Services Organisation that maintains and enhances the HSC Regional Data Warehouse, using a range of technologies. She also has responsibility for Integration services and the development of small-scale bespoke regional web applications for the HSC.
GOALS (objectives this person hopes to achieve through MIDAS)	The Regional Data Warehouse is mainly used by trained Information staff. Susan would like to make the data accessible to decision makers in the course of their day-to-day work.
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	Buying into expensive proprietary toolsets which require high level of training, knowledge and expertise.
TECHNOLOGY SKILLS (% level)	
IT and Internet	80%
Software	80%
Mobile Apps	50%
Social Networks	30%
DATA ANALYTICS SKILLS (% level)	

Integration	60%
Analytics	40%
Visualization	40%
Interpretation	20%

2.3.2 Scenarios and User Stories

The term "Looked After Children" refers to a diverse group that varies in terms of age, ethnicity, the reason for being looked after, age of first entry into care and duration within care. However, it is fair to say that, while some children and young people in care can go on to enjoy success, as a group, educational and other outcomes tend to fall significantly below those of the general population. Such large shortfalls are not just concerning in themselves, but also as predictors of later life chances. As a group, looked after children are at far greater risk of experiencing social exclusion.

The key outcomes would include being able to track patterns of behaviour over time; visualisation of data on movement of children in and out of different types of care; tracking Adverse Childhood Experiences into the future to be able to identify how we address their needs (and whether this was effective); examination of trends in data will help HSC professionals to enquire more deeply (policy informing practice).

Actions emanating from the draft Looked After Children Strategy will be prioritised. In particular, the ongoing Review of Secure Care and its interface with other regional specialist children's services, will enable the review team to track the cohort of children who frequently are referred to these services and help to identify patterns of behaviour. This will help to inform and consolidate the Review's findings and in the longer term, allow for the exploration of variation in HSC practices with regard to this group of children and to identify lessons to help inform policy development and service design.

2.3.3 Visual Requirements

Due to limited experience with the joint data approaches, it is quite difficult to specify what visualisations would be the most valuable to investing in this pilot. The targeted organization has not yet real understanding about the strength and possibilities of the joined-up data and some of the included datasets are completely new for the end users in their perspective. The needed understanding for the final needs to visualize the data and results will build up over time and it may be that through that process some outputs could get solidified as routine via a dashboard. The strength of MIDAS is having the linked datasets, but that strength brings complexity that will require considerable understanding – which the pilot do not yet have.

However, there are a number of items worth reiterating that it is thought to be essential to have as part of an output toolbox:

- On the final output dashboard having a list of the datasets/universes from which the variables were extracted
- The datasets/universes from which the variables were extracted is part of the variable name
- The ability to have a list of favourite variables
- The ability to save a query that has been built
- The ability to save and export any visualizations that have been generated
- The ability to save and export any underlying data used to generate an output.
- Ability to load additional datasets as they become available (e.g. Educational attainment).
- Create a printable report from the selected dashboard and visualisations

In addition, the visualizations would require at least:

- Use of mapping with the ability to focus on areas according to Northern Ireland Statistics and Research Agency (NISRA) geography
- Dynamic cross-filtering figures of reports and results of the analytics.
- It would be interesting to visualise the increase in Children In Care and Children on the Child Protection Register since 1995 and any area differences (E.g. by MDM or Trust etc.) taking into account the child population in the areas.

2.4 Republic of Ireland

The “Healthy Ireland” framework and related policies is a key delivery objective for the Republic of Ireland with a key focus on the diabetes challenge. Chronic disease management poses an issue for patients with one or multiple disorders. Diabetes is a fast-growing disease affecting 10% of the population of Ireland and has significant symptoms for patients and life changing factors. It is estimated that there are 190,000 people with diabetes in Ireland (Institute of Public Health in Ireland, 2007)[9]. Approximately 30,000 (15%) of these people do not have Type 2 diabetes, but either have Type 1 diabetes, or genetic or secondary causes of diabetes. The remaining 160,000 (85%) patients have Type 2 diabetes. A significant proportion of these patients (20-30%) remain undiagnosed. It is expected that the number of people with Type 2 diabetes will increase by 60% over the next 10-15 years. The challenge is to gain insights from existing data to make policy decisions for the maximal impact on the wellbeing of these people and to drive operational efficiency.

9 https://www.publichealth.ie/sites/default/files/documents/files/Making%20Diabetes%20Count%20What%20does%20the%20future%20hold.pdf

The MIDAS project proposes to afford meaningful information insights that can enhance policy. In essence, the success of MIDAS for the Republic of Ireland (ROI) stakeholder community will be determined by its ability to provide insights from existing data to make policy decisions for the maximal impact on the wellbeing of these people and also to drive operational efficiency.

There is no single centralised diabetes register for Ireland. Sharing and merging of appropriate data sets from various sources is an issue. This involves merging the datasets on a single dashboard to provide a picture of diabetes in Ireland. Completing the picture with accurate and up to date data for policy and decision making is the goal and would enable moving from treatment costs to prevention costs. Demographic insights, age, gender and socio-economic dimensions are also central to the analytical process. Information collated will be utilised to determine additional resources required to drive improvements in care delivery.

The key stakeholders are the Department of Health (DoH) for policy planning and many divisions of the Health Service Executive (HSE). The purpose is to plan a platform which supports policy decision making from a DoH level to public health and HSE Governance level, particularly in resource management of budgetary allocation and staff planning and allocation, analytical insights to support decision making and efficient healthcare delivery from a perspective of caseload managers.

2.4.1 Stakeholders

The Republic of Ireland are reviewing methods of ensuring a healthier population in general. Their focus is currently on persons with diabetes which can significantly affect the health and wellbeing of the person. The Republic of Ireland has not engaged in wider workshops or engagement with stakeholders or Policy Board until the deliverables such as the dashboard prototype is at a more advanced level. Nominated users participated in the user evaluation and testing of the first iteration of the dashboard type based on the planned user of the platform in the future. It is planned to have a wider engagement in the coming months based on the first user experiences and feedback to ensure the dashboard prototype is fit for purpose for the users defined.

Personas of stakeholders who were identified to participate in iterative testing and evaluation of initial prototype.

Senior Manager with IT Focus on Systems Availability to support Care

NAME	Pamela

AGE	48
JOB ROLE	Programme Lead, CCIO Management Team, Office of the Chief Information Officer, Dublin
LOCATION	Dublin
TRAITS	Inquisitive, experimental re IT,
BIO	Education – MSc in Health Informatics. She has worked in Health IT realm for over 10 years. She has lead developments in data quality and audit systems. Her particular interest is in making data useful through automated collection to reduce workload of audit data collection. She is particularly interested in reports and use of the data to support decision making.
GOALS (objectives this person hopes to achieve through MIDAS)	Pamela is keen to increase the use of data available across multiple layers of an organisation
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	Data excel sheets and the pain of trying to decipher information from rows and rows of data. It would be excellent to have visualisations of data which can support the decisions needed.
TECHNOLOGY SKILLS (% level)	Pamela have excellent skills in use of IT systems, email and communications systems and use of internet to seek information
IT and Internet	80%
Software	70%
Mobile Apps	60%
Social Networks	40%

DATA ANALYTICS SKILLS (% level)	Pamela has skills in use of data analytics tools such as Tableau, Qlik and is familiar with dashboard use
Integration	60%
Analytics	80%
Visualization	80%
Interpretation	70%

Senior General Manager, Strategy and Planning

NAME	Mary McCarthy
AGE	42
JOB ROLE	General Manager
LOCATION	Dr. Steevens Hospital, Dublin
TRAITS	Organised, senior decision maker
BIO	Education: BSc Business Administration, MSc in Strategic Management Mary is a strategic planner associated with the Office of Strategy and Planning. She previously worked as a manager in an acute hospital in the Republic of Ireland. Her role entails ensuring access to information to support evidence based decision making and resource allocation.

GOALS (objectives this person hopes to achieve through MIDAS)	Mary is keen to have quick and easy access to data from numerous sources which will provide a detailed picture of the state of a situation. For diabetes, this would mean accurate figures of incidences and distribution and treatment points in order to establish where resources need to be dispatched
FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	Numerous sources of data, no one truth. Lack of analytics tools in a simple visualisation to support decisions.
TECHNOLOGY SKILLS (% level)	*Competent in Administration systems and some clinical systems associated with clinical care delivering, administration and planning*
IT and Internet	80% (Good access and skill)
Software	60% (Good access to clinical and admin systems)
Mobile Apps	30% (Not routinely used)
Social Networks	40% (Professional networking to an average level)
DATA ANALYTICS SKILLS (% level)	*Is competent in analysing complex scenarios and determining key aspects of reports and findings*
Integration	30% (Aware of silo aspect of many excellent sources of information, would really benefit from greater access and integration of data sources)
Analytics	60% (Is competent in analytics systems but does not always have the correct tools for the scenario or information source to hand
Visualization	50% (Can interpret data when presented effectively so is keen to access tools which present visualisations well)

Interpretation	80% (Has strong skills in decision making and interpretation)

Clinician Persona of User Evaluator

NAME	Dr John Moriarty
AGE	46
JOB ROLE	General Practitioner (GP)
LOCATION	Co. Louth, Ireland
TRAITS	Communicator,
BIO	Education: Medical Degree, Irish College of General Practitioners (ICGP) and Fellowship of GP, Ireland. GP Medical Practice would have a population of 6000 with 10% in the over 65 age bracket. Focus for the practice would be on general population health with specialist clinics for diabetes, asthma, dermatology, blood pressure and cardiac monitoring and general health screening. As a GP, he has access to internal medical records for patients registered with the practice, medication listings and laboratory reports for tests ordered from the practice. Wider integrated access to summary information would be ideal for diagnostic or planning care such as in patient episodes of care in acute hospitals.
GOALS (objectives this person hopes to achieve through MIDAS)	Optimise patient care through having the right information available at the time of the visit and access to appropriate information to support diagnosis Improve communication channels for all interested parties in all directions Improve access to information for the patient to support self care

FRUSTRATIONS (the pain points they'd like to avoid through MIDAS)	Lack of Integration of data across healthcare providers Greater visibility of the wider evidence relating to chronic disease management such as diabetes with a national picture Require greater shared policies for care following patient journey
TECHNOLOGY SKILLS (% level)	*Competent in GP Systems and Systems in general*
IT and Internet	80% (Good access and skill)
Software	80% (Good access to GP Software)
Mobile Apps	20% (Not routinely used but used when appropriate)
Social Networks	60% (Strong part of medical practice role)
DATA ANALYTICS SKILLS (% level)	*Has strong capacity to absorb information provided in a form which supports analytics review*
Integration	70% (Of data available in GP System, good integration but would benefit from external sources of information integration)
Analytics	70% (GP System supports good use of analytics)
Visualization	50% (System would benefit from more modern approach to visual display
Interpretation	70% (Supports decision making and interpretation)

2.4.2 Scenarios and User Stories

Forecast Tools for the Republic of Ireland scenarios are critical to shape policy within the Dept of Health and at policy level within the HSE. We wish to facilitate forecasting of a national picture of diabetes across Ireland, geographical area, urban and rural and by other correlation trends such as deprivation patterns and medication trends. Predictive modelling to support policy makers to plan diabetes programmes based on preventative measures rather than treatment options is a core component. Caseload management of persons with diabetes from an efficient staffing perspective as well as healthy lifestyle for individuals is key.

2.4.2.1 Scenario

The Republic of Ireland (ROI) scenario is based on access for policy makers to a national diabetes register with up to date information on all persons with diabetes in their region. The aim is to support self care for patients with diabetes through clinical health care professionals sharing information relating to diabetes care in real time providing supported decision making for the person with diabetes.

Higher level policy makers need information from a range of sources amalgamated in a single workspace or dashboard to support decision making and require quality information which is evidenced based and supported by quality standards which gives valuable information to determine the level of risk in a community, in the hospital on any given day or period.

Issues in the ROI are the lack of:

- Access to health records at the point of care or lack of sharing of information across hospital/community boundary
- IT software to facilitate information recording in a health record which is accessible for others
- Database to support caseload management (prioritisation by risk and health needs)
- Statistical and trend analysis of care episodes to determine risks
- Access to standardised quality data, following set criteria for collection and validation
- Real time data which can be comparatively analysed and utilised and trends identified

2.4.2.2 User Stories

2.4.2.2.1 Regional Story

Visualisation and analytics of the data with trend analysis would offer core clinical information on the validity of treatments and education plans. It would also assist in caseload management to place resources such as clinical nurse specialists and endocrinologists in areas which have a high density of persons with diabetes.

2.4.2.2.2 National Story

Nationally there are significant data available from national data sets such as HIPE and PCRS but in isolation, this does not give a true picture. Data modelling can be activated to provide for resource planning and caseload management for persons with diabetes. At a national level, financial resources can be mapped according to population trends and incidences and at a regional level, decision support can be offered to healthcare professionals to plan health wellness of targeted populations.

2.4.3 Visual Requirements

The most important focus is a single access point for the Republic of Ireland to view information on the wider diabetes situation nationally. It can then support analytics and visualisations which enable data driven policy development and resource allocation based on population and trend analyses from evidence based data.

Data visualisation takes many shapes in the Republic of Ireland. From a national perspective, Health Atlas[10] is a resource used which has open source data geomapped by hospital regions and community regions. Health Atlas Ireland is guided by Health Intelligence/Knowledge Management HSE in collaboration with numerous bodies. There are also numerous dashboards available to senior decision makers in individual hospitals but few national dashboards which give a comprehensive overview across the Republic of Ireland. As the personas in the ROI ranged from High level users to clinicians who could be making strategic decisions based on information displayed, it was difficult to perceive how one dashboard would be able to accommodate all needs. This was reflected in the user evaluation. Some users were adept at using statistical tools and predictive modelling whilst others had no experience of this.

The Irish pilot has a broad target audience and accordingly all common formats of visualisation should be provided to satisfy user requirements. The following data visualisations are identified as a priority:

- Data visualisation directories geomapped by hospital and community regions.
- **Choropleth Map** of the Republic of Ireland depicting:
    - Population distribution across Ireland
    - Deprivation patterns geographically urban and rural
- **Dot Density Map** of Hospital showing distribution of
    - Acute Care services
    - Primary Care
    - GP and Health Centres
- **Correlation charts** with **scatter plot** to show incidences of diabetes by employment rates and deprivation
- **Population Pyramid** to show distribution of diabetes Type 1 on left side and Type 2 on right side by age, region, comorbidity

10 https://www.healthatlasireland.ie/

- **Bubble graph** to show variables as above (age, region, comorbidity) with diabetes in a different format
- **Cumulative Curve** to show incidences of diabetes over a time period, e.g. a 10 year time span.
- Data Modelling and predictive analysis with forecasting tools to predict diabetes incidences, short term and in to the future
- **Percentile Charts** and run charts and Ordered Bar Chart to show budgetary allocation of resources (finance allocation and staffing location (Clinical Nurse Specialists)) related to diabetes compared to incidences of diabetes diagnosed for that region.
- Ranked lists of comorbidities/prescribing patterns per geographical regions
- **Stacked Columns** to show adverse events aligned to medication management of diabetes
- Chart depicting symptoms of diabetes relating to medication used to determine efficiencies of new medications methods
- Process Control chart to demonstrate that hospitals across Ireland are meeting diabetes care targets (Care of diabetic foot, retinal screening etc.)
- **Line Chart** depicting the number and frequency of hospital admissions per type of diabetes or complication of diabetes (amputation, vision issue, hypo or hyper situations).

3 MIDAS Dashboard System Architecture V2

The MIDAS Dashboard V2 conceptual structure is similar to the one presented in D5.1 for V1. The only change is the addition of GYDRA tool to the figure. The MIDAS Dashboard is built with two main components and includes connectivity to external resources which are shared by all MIDAS Pilots. The main components of the MIDAS Dashboard are the user interface (UI) and middleware component. The UI delivers the modular dashboard generator and the middleware serves connectivity to the UI for all necessary services flexibly. Figure 3.1 describes the components used to create the MIDAS visual analytics tools and this chapter describes the main approaches used to implement these on the second iteration of the MIDAS prototype.

3.1 User Authentication

The MIDAS Platform includes several components from several network domains. The core analytics components are located in closed intranets in every Pilot instance. The news, social media and MEDLINE analytics instead are located on external web resources and shared with every Pilot. To keep the UIs of MIDAS Platform uniform they all should work with single sign in within a Pilot. In practice, this means that user is requested to sign in only once and then all external and internal services are utilizing the same authentication and authorization service.

In MIDAS V1 the authentication and authorization of end user services are archived by using a separated authentication server located at MS Azure cloud which is accessible from every partner and pilot. Due to technical complexity of the system maintenance, the authentication and authorization systems are planned to be moved to MS Azure AD B2C service based system, where Azure maintenance takes care of the backend system updates and MIDAS partners are concentrating to the secure implementation of MIDAS Platform.

All solutions are implemented using OAuth 2.0 protocol. This is utilized on all subsystems of MIDAS Dashboard to ensure the single user security on the system. In MIDAS Dashboard this means that end users are authenticated through the system when they use UI. This authentication creates a session token for the system which is then used with all data access requests from UI to OpenVA. The token is valid only certain time (after user activity has ended) or until user signs out from the system.

To achieve the single sign on experience the same authentication token and service must also be used by the external services like Social media dashboard and news and MEDLINE analytics. These will be implemented after Azure AD B2C authentication is found functional and taken into use at MIDAS Dashboard first.

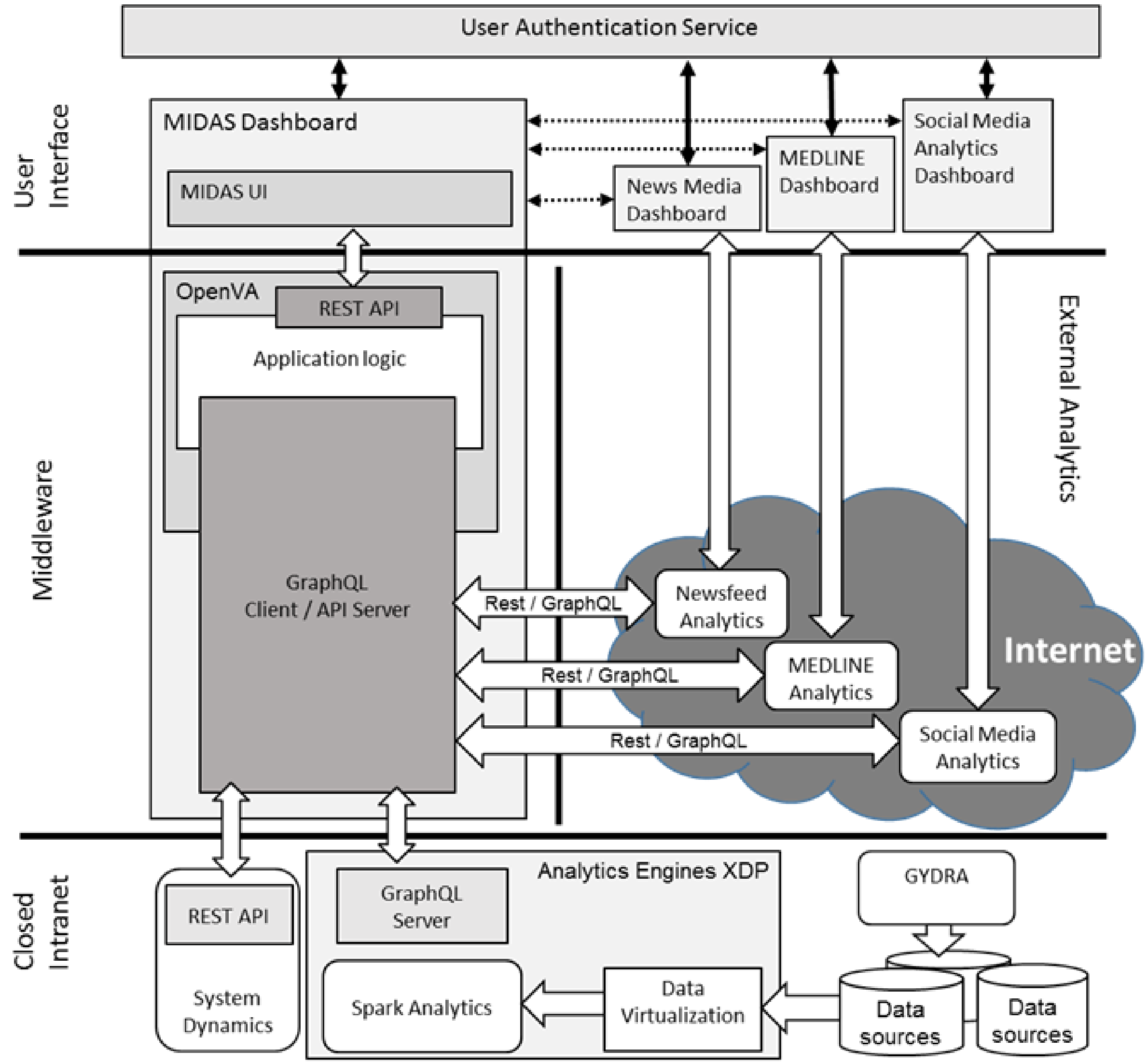

Figure 3.1: MIDAS Dashboard concept. The MIDAS Dashboard includes two main components: The MIDAS UI and OpenVA based Middleware component. The UI is communicating with the end user and the OpenVA is responsible for communication between UI and all analytics services used in MIDAS concept. System dynamics server component is postponed due to technical issues.

At MIDAS we are using an external service - Event Registry - for which we obtained a license for the use in MIDAS, and in which the news dedicated dashboard is based on. We are currently discussing with the system owner whether the single sign on would be a viable option for them, and how can it be integrated.

3.2 Technical Interface Components

The middleware technology in MIDAS Dashboard has two main responsibilities:

- Handling the metadata to enable smart logic for UI (OpenVA logic)
- Apply efficient communication between UI and different analytics (GraphQL)

3.2.1 OpenVA 2.0

OpenVA[11] acts as a gateway between the Midas UI and different other services like Simantics System Dynamics server[12] and MIDAS analytics layer. It enables REST/GraphQL interface for MIDAS UI to communicate with and REST/GraphQL for communication towards analytics systems. This enables feasible uniform access for MIDAS UI to different kind of data sources and services but also implementing other kinds of MIDAS UI in the future like native UIs for some dedicated mobile environments.

The OpenVA is developed by VTT with Spring Framework[13] using the following projects:

- Spring Boot to allow quick development on production ready application.
- Spring Data to internal data access of metadata.
- Spring Security for protecting the application with authentication and authorization.

Developing with Spring is done with Java which also allows native utilization of JavaScript with Rhino JavaScript Engine if needed. Handling metadata with OpenVA brings additional value for utilizing the other services. E.g. models of System Dynamics contain several values but only some of them are in general modified. As a result OpenVA can handle storing the different views and access to the model independently of the model execution.

OpenVA is handling the metadata information with a specific metadata model, which is updated according to the available data, analytics and visualisations on the environment. In the MIDAS pilots the OpenVA metadata model is Pilot specific and during the development of the system the options to update the metadata model automatically are studied. A reference of a typical OpenVA metadata model was described in D5.1. The OpenVA component also persists the state of the dashboards defined in MIDAS UI and the database schema for these are presented in Appendix 1.

3.3 User Interface Components

There are five systems with separated end user interfaces in the MIDAS Platform:

- MIDAS Dashboard
    - The main UI in MIDAS Platform to generate decision support dashboards for Policy Makers
- GYDRA data quality management tool

11 https://www.thinkmind.org/index.php?view=article&articleid=infocomp_2013_3_10_10030

12 http://sysdyn.simantics.org/

13 https://spring.io/projects

    - Data quality management tool for MIDAS data analytics
- Social Media Platform
    - System to manage and analyse in detail the Twitter campaigns
- NewsFeed analytics management Dashboard
    - Dedicated dashboard based on the Event Registry technology for news analytics
- MEDLINE analytics management Dashboard
    - Dedicated dashboard based on elasticSearch and Kibana technologies for detailed MEDLINE abstract analytics

The MIDAS dashboard is the main user interface of the system and the three other dashboards are serving the management interface for the complex analytics needed with the Social Media, Online News and MEDLINE resources. The GYDRA web tool enables users to correct and analyze raw data and estimate data quality. The results of the other systems can be linked to the MIDAS Dashboard with widget components. The common user authentication service is utilized by all four dashboards.

3.3.1 MIDAS Dashboard

The user interfaces of the MIDAS Dashboard are aimed to support reasoning loop workflow. The concept of reasoning loop support is visualized on Figure 3.2 below. The main idea on reasoning loop workflow was described in deliverable D5.1. The main concept with the reasoning loop approach is that a user can easily modify and fine tune the applied analysis and visualisations based on investigation of the previous round. This creates the basic requirements for the widget functionality.

The reasoning loops were reorganized after the first end user comments and Figure 3.2 reflects the latest approach. The main change is that user first selects the visualisation together with analytics and only after those do they select the desired data and variables. The basic principle still stays the same and the system shows only possible options to the user in each step of selections.

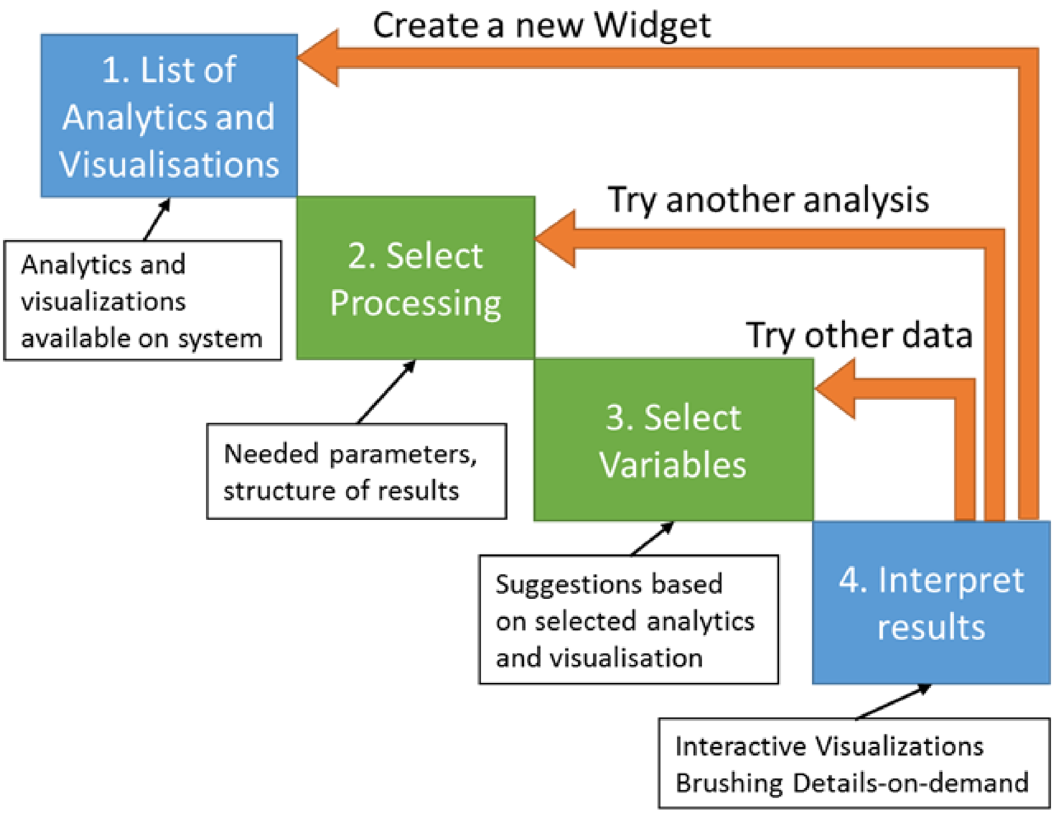

Figure 3.2: Reasoning loop logic v2. The MIDAS Dashboard will support end users to use reasoning loop logic when preparing Dashboards and investigating data. The key concept with reasoning loops is to possibility to modify the earlier selections and create new analytics according to the detections from them.

3.3.1.1 UI Components

Workspace

The main component of MIDAS Dashboard is the Workspace, where a user creates the Dashboards with widgets. In practice the workspace is a virtual board where the user can move resize and set widgets in a preferred order. When needed the saved workspaces, dashboards, can be shared to other MIDAS Dashboard users. Main functionalities of workspace are save, save as, open, undo action, name the workspace and share the saved workspace Figure 3.3 demonstrates the workspace and menu of the MIDAS Dashboard

Widget Wizard

New elements for the workspace are added via element specific wizards. All available wizards in MIDAS Dashboard deployment are listed in widget menu of MIDAS Dashboard. The version 1 of MIDAS Dashboard supports three types of widgets where two of them have a dedicated wizard to create the widget.

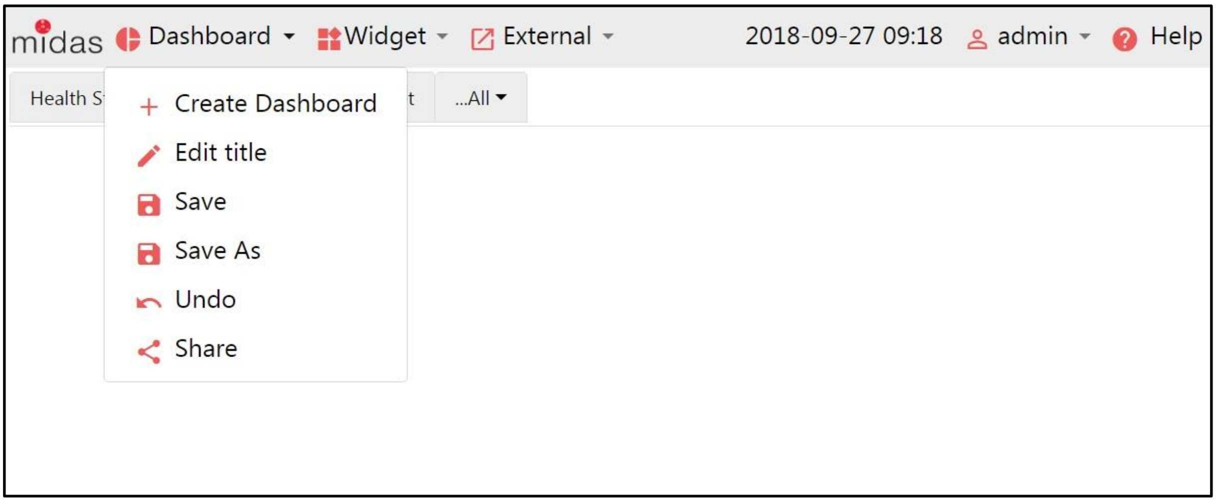

Figure 3.3: The workspace and menu of the MIDAS Dashboard.

When opening the wizard it appears on top of the workspace and has some number of steps to generate the widget. For example the wizard of social media widget includes only a single step where the user selects the social media campaign containing the main results which are wanted in the dashboard. When selecting “Add Widget”, the system requests data from social media platform and generates a widget in the workspace. Figure 3.4 shows an example of the social media widget wizard.

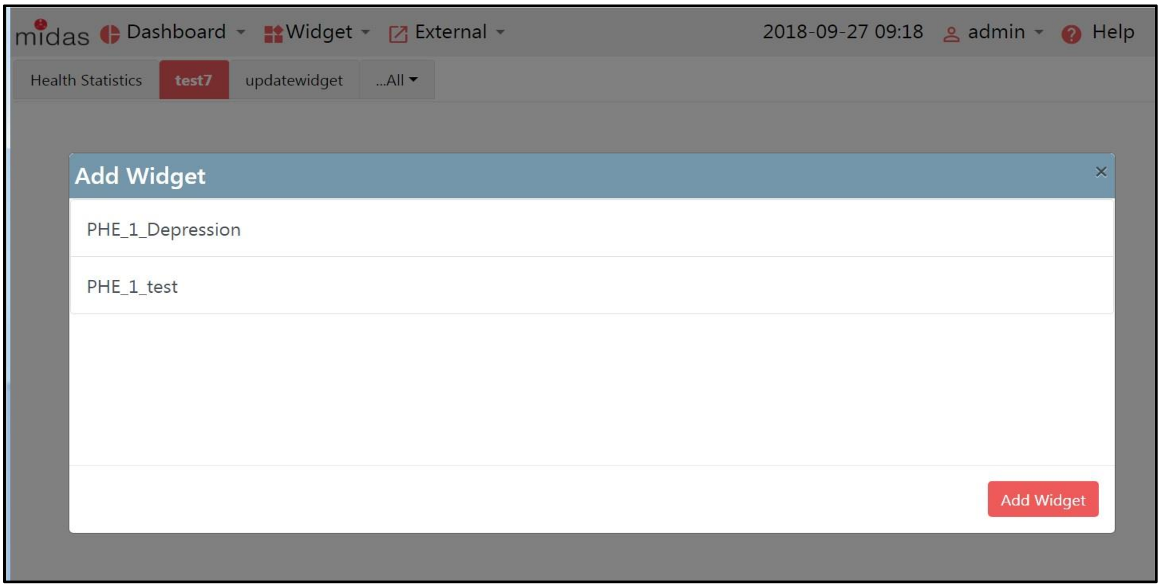

Figure 3.4: The simple widget wizard of the MIDAS Dashboard for Social Media widget.

The main analytics widgets have a more complex wizard. On the first step the widget requires user to select the wanted visualisation and analysis type and on the second step user must select the wanted data for the analysis. Again the widget is generated by the system when “Add Widget” button is pushed. Figure 3.5 shows an examples of the two steps of general widget wizard.

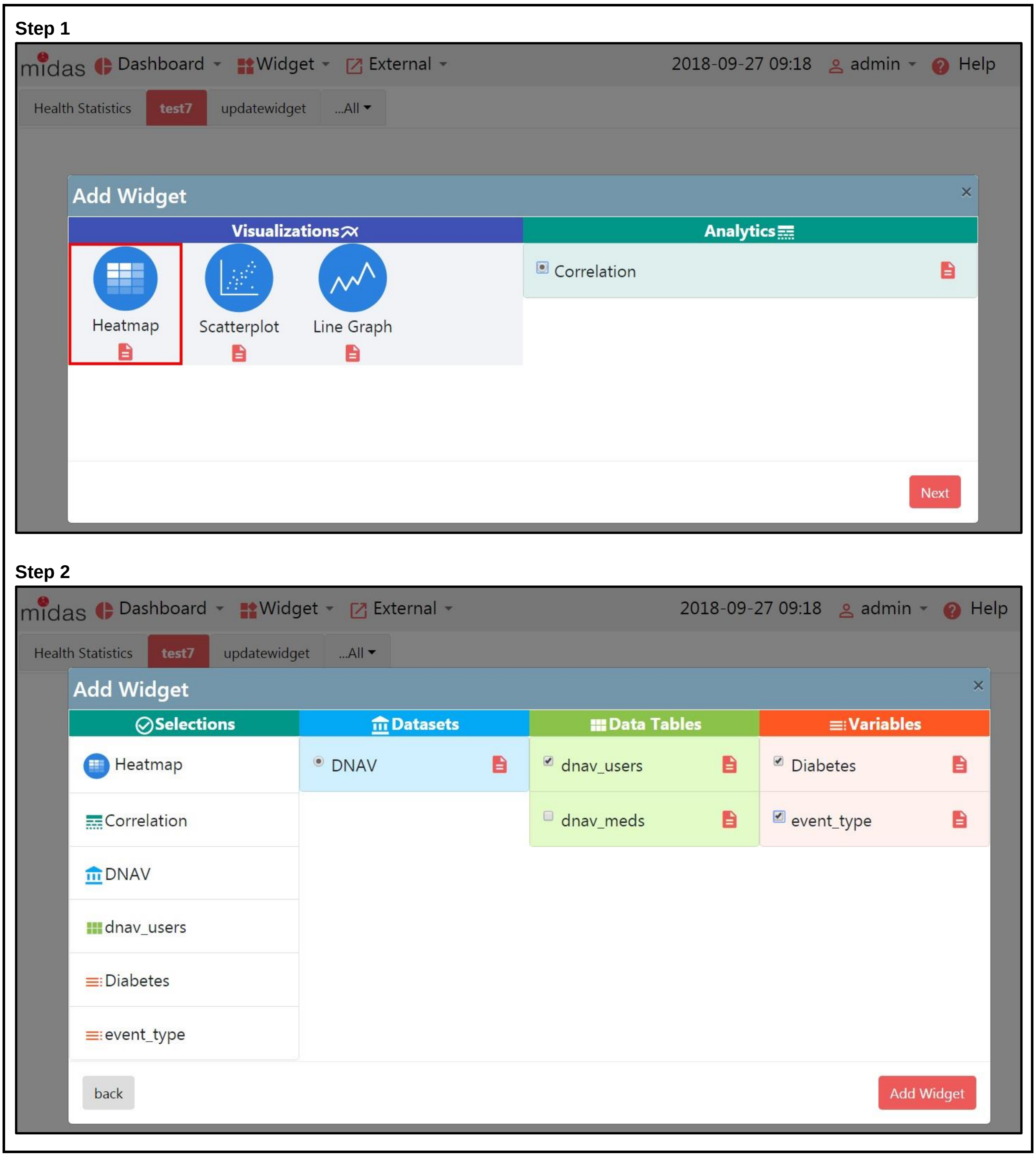

Figure 3.5: The widget wizard of the MIDAS Dashboard for common analytics and visualisations. The step 1 selects the visualisation and analytics with it, the step 2 selects the data which is feeded for the analytics.

Widgets

Widgets are the key components of the MIDAS Dashboard. With the visualisations presented in the widgets, MIDAS Dashboard connects external and closed intranet analytics into a single service where a user can investigate at the same time results of analytics from the local data or external sources like a Twitter social media campaign. All widgets have the same base components. They appear in the workspace like windows which can be resized and relocated on the dashboard. They also include an option to add text comments to them through a comment button and remove them by removing cross at top right corner. The rest of the functionality of a widget depends on the widget type. Most of the figures support hovering by pointer to get more details out of them and some figures allow adjustments to change for example time scale or zoom into certain values. General definition of custom widgets on MIDAS Dashboard is presented in Appendix 2. Examples of different widgets from V1 are presented in Figure 3.6.

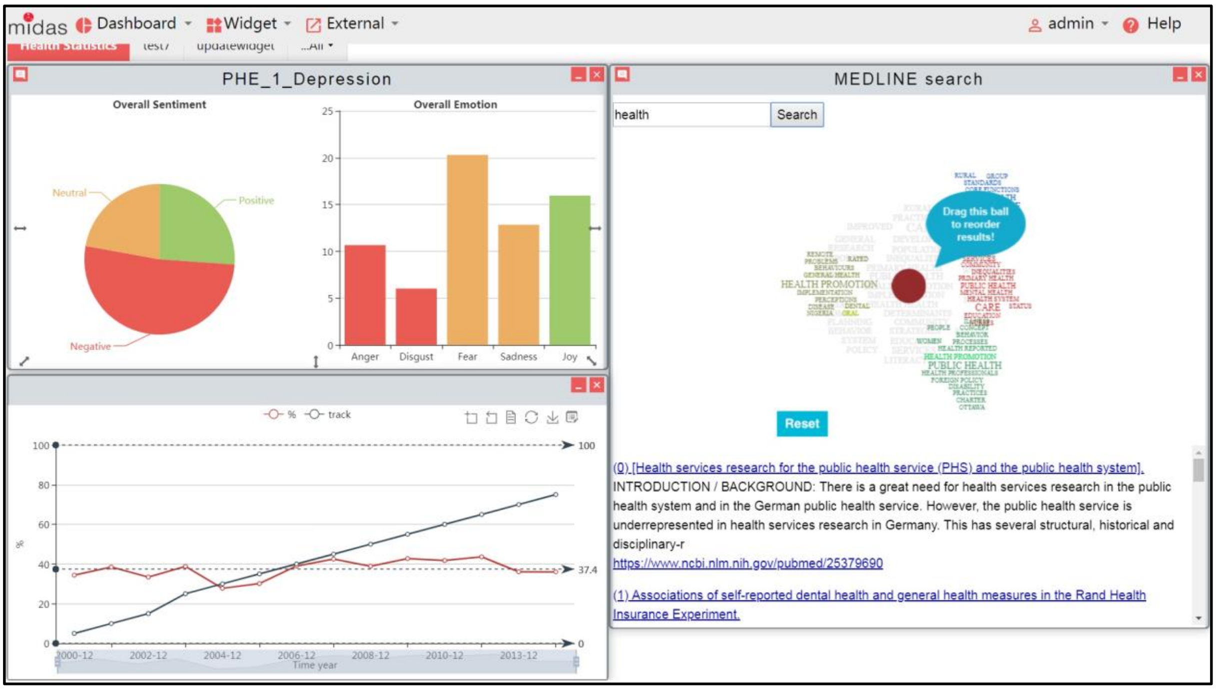

Figure 3.6: The workspace of the MIDAS Dashboard with three widgets on it: Social media at top left, Common analytics with line plot at bottom left and MEDLINE widget at right.

Common Widget

This widget type is where the general analytics with local data are visualized. The widget will support various visualisations where at least Line graph, Bar graph, Scatter plot, Heat map and Choropleth Maps are supported. You can see the example of the widget in Figure 3.6 bottom left corner.

Pilot Specific Widgets

Each pilot will have also pilot specific analytics and those are developed separately for each pilot and have dedicated visualisations or combination of visualisations available. These are drawn to MIDAS Dashboard via Pilot Specific Widgets. Development of these widgets are done in close collaboration with Policy Board UX test participants.

Social Media Widget

This widget draws the main results from the selected social media campaign of the social media platform of MIDAS project. It includes two figures, a pie chart and bar plot which represents the mood and feeling of people who have answered to the campaign. In the future iterations, options to allow user to select which figures from the social media campaign are presented will be studied. You can see the example of the widget in Figure 3.6 top left corner of workspace.

MEDLINE Widget

This widget provides advanced MEDLINE content based on the SearchPoint technology, enabling advanced search from MEDLINE/PubMed abstracts. It does not have any wizard but the widget itself is fully interactive. The user types the search keywords to the box on top of widgets and the widget requests the results from the external search engine. As a result, the user gets a wordcloud where he can change the weights of search ranking by moving the red dot over the word cloud. It enables end users to bring the scientific knowledge directly as a part of the Dashboard they are building/using. You can see the example of the widget in Figure 3.7 and at right in Figure 3.6.

News Analytic Widget

This widget will provide insights in the news monitoring information of a certain topic extracted from the data provided by the EventRegistry available APIs that MIDAS as access to. It will be integrated in the prototype V2, enabling some iteration with the user, permitting the visualization of the available stream of news adapted to the topic in focus. This will complement the information in the main MIDAS dashboard provided by the analytics of proprietary data, the MEDLINE biomedical information from scientific articles, and the social media campaign analytics.

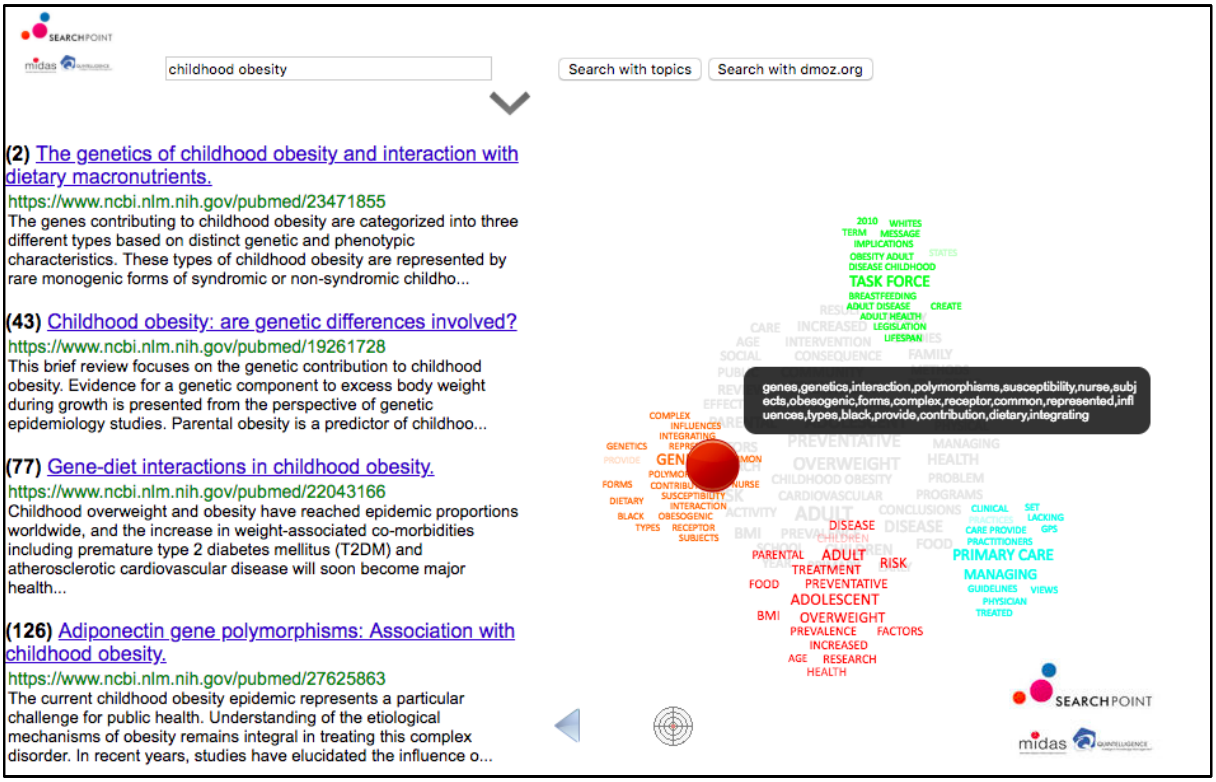

Figure 3.7: Screenshot of the MIDAS MEDLINE custom widget after a query on “childhood obesity”, exhibiting several clustered keywords representing areas of interest that the user can focus on by moving to it the red searchpoint which results in reindexing of the list of related items. Notice that in our example moving the red searchpoint to GENETIC resulted in positioning among the top 5 an article that otherwise would occupy position 126.

3.3.1.2 Technology

The second prototype of the MIDAS dashboard is developed by utilizing the basic web-technologies, which includes HTML5 (Hypertext Markup Language), CSS (Cascading Style Sheets) and JavaScript. Moreover, to speed up the process of development and to provide latest UI functionalities, following pre-build libraries and frameworks are used.

- **jQuery**: (version 3.3.1) is a JavaScript library that allows easy and fast manipulation of the DOM (Document Object Model) elements in the HTML document. It also provides functionality to interact with external REST APIs by using AJAX (Asynchronous JavaScript And XML) and handles various other important tasks required to develop a web-application.
- **Bootstrap**: (version 4.1.3) is a CSS framework for designing and development of modern and responsive front-end of the web applications.
- **Gridstack.js**: (version 0.4.0) is a jQuery based plugin that allows developing responsive widget based layouts. It is also compatible with the bootstrap framework.

- **eCharts**: (version 4.2) is a JavaScript based library for rendering various types of graphs by utilizing a HTML5 canvas.

Most of the libraries and frameworks described above are the same, which were used for the development of the first prototype. However, older versions are replaced with the latest releases. Moreover, the older chart library flotr2 has been replaced with a new library eCharts. The eCharts library is more robust, and it can renders a large amount of data points with much less delay as compared to the older library. In addition, the new chart library provides vast amount chart types to choose from, which allows us to add more visualizations options into the MIDAS dashboard UI.

3.3.2 Social Media Dashboard

IBM is currently using Twitter to provide insights about the public's sentiment towards health policies. In the previous iterations we planned to include other sources, for example, Facebook, Linkedin, Google+, Instagram, Reddit but after GDPR concerns and approval issues we decided to not directly implement them instead we decided to focus on the implementation and architecture of the Social Media Dashboard. We want to perfect the user and dashboard experience and we think we can do that with one source (Twitter). Instead of focusing on specific social media platforms, we want the Dashboard to be easily extendable via plugins or data sources. That way if a certain policy maker in a country wants to include a different social media platform, they can write adapters/plugins for the Dashboard. However, depending on Policy makers needs we may decide to implement another one or two sources depending on popularity and terms of service.

IBM's Twitter chatbot will ask members of the public a series of questions about the given health policy. The entire response object from Twitter will be stored and augmented further with annotations from IBM's Watson services on IBM Cloud. An example of the Twitter direct message object can be found from their webpage. Watson will then add sentiment and emotion values to the object, perform named entity extraction, concept extraction, identify people names and identify place names.

3.3.2.1 UI Components

Authentication – Currently the dashboard has its own user authentication, allowing registration and login to all users. For future iterations there is a plan to have a common authentication system across all the dashboards, thereby simplifying access controls and making easier for dashboards to integrate with each other.

Ability to run multiple campaigns at once – The dashboard is capable of running multiple campaigns at once from the same twitter account. You can imagine a policy maker wanting to know the public's opinion on variety of issues at the same time this feature enables that.

Displaying the results of campaigns questions - In order to deliver quantitative results, similar to those seen in online survey tools, such as survey monkey, the dashboard displays charts showing the results of each question. This means highlighting how many “Yes”, “No”, “Maybe” or “I don't know” responses were given as responses by the public and also how the public answer multiple choice questions.

Display a small number of deeper insights - The dashboard displays a small number of deeper insights. These include most common suggestions that the public have made for the given policy, gathering and displaying responses to multiple choice questions that were not included in the options given to the public. These types of results will be more valuable to the policy makers as they will help to identify issues and concerns that the public have with the policy.

Ability to create a campaign & Chatbot through a page on the dashboard – Previously, the campaign details, chatbot and questions logic were controlled by configuration file which was very difficult for policy makers to use without support. We have a built a ‘ChatBOT creation’ form into the Social Media Dashboard where we have made it easy for Policy makers to create their own campaigns and chatbots without worrying about configuration files. Policy makers can customize the name, questions and even give answers to commonly user asked questions so the chatbot can respond with custom answers for each campaign. In future iterations we plan to make this ‘ChatBOT creation form’ more powerful, by allowing policy makers to control the flow of the conversation and allow them to customize what happens when user answer specific answers to questions. e.g: If a user answers ‘no’ to this question ask ‘why’.

Ability to edit a running campaign – While a campaign is running, the system may surface information that the policy maker will wish to address. For example, the chatbot might highlight that people keep asking about what certain options mean in multiple choice, that it can't provide an answer to. At this point the policy maker is able to edit their campaign via the Edit campaign form to provide answers to these questions, so that survey/campaign is better.

3.2.2.2 Technology

For the frontend, IBM is using following stack:

- **Angular: (version 4)** is a framework developed by Google for building single page web applications which include server side rendering for increased performance. Google skipped version 3 and released version 4, which the dashboard has been upgraded to.

- **Bootstrap: (version 4)** is a CSS framework for designing and development of modern and responsive front-end of the web applications.
- **Webpack: (version 3)** is a tool used to compress, minify and aggregate javascript, css and assets reduce the size and load time for opening the application.
- **Highcharts: (version 0.5)** is a Javascript library used to draw custom charts, visualisations and animations.

The backend will consist of:

- **Node.js: (version 8)** is an open source Javascript framework used to develop scalable server side / cloud based applications.
- **Express: (version 4)** is an open source Node.js module used for building REST based API's. IBM will investigate using GraphQL in later iterations.
- **Cloudant:** Cloudant is IBM's hosted version of the open source technology CouchDB, with additions to make it more suitable for big data solutions. Cloudant is a NoSQL based used to store JSON documents and run complex searches with Apache Lucene.
- **REDIS:** is an open source in memory database designed to store small amounts of data, but with an extremely quick response time. It is frequently used for session token or state management.

3.3.3 News Media dashboard

This dedicated dashboard enables the monitoring of worldwide news over 32 languages (including Basque and Finnish, covering fully the MIDAS use-cases). Based on the common MIDAS account for the usage of Event Registry technology, QUIN is developing several Public Health focused tools based on state-of-the-art text mining technology. Those are fed by the open data sources of (i) digital multilingual worldwide news, and (ii) integrating the MEDLINE knowledge base insights.
The visualisation components provided by a query over a health related topic provide the user with a wide range of visual exploration modules over the insights provided by the news stream related to the topic in focus. In particular, it clusters news articles covering the same event, permitting the user to analyse the storyline of the happening, from the first cases throughout its development (which is very useful, e.g., when analysing the response to an epidemics).

The common challenge of using free text documents (such as news articles and official documents) to support public health decision making is addressed by a classifier based on the MeSH Headings categories (recently made available at for demo at http://midas.quintelligence.com) that will serve us to analyse any kind of texts and provide it with relevant MEDLINE-related metadata. The metadata attached to text documents describes the document content in a uniform and standardised way. For the above mentioned sources such metadata consists of geographical locations, types of events, health-related categories, source identification and timestamps.

As described in the deliverable D3.1, Section 3.2, QUIN provides access to newsfeed.ijs.si through an agreement with Jozef Stefan Institute (JSI), Ljubljana. Newsfeed provides a real time stream of annotated, multilingual news. This stream is being further annotated with MeSH with modules developed in WP3 and WP4.

3.3.3.1 UI Components

The UI components of this dashboard enable a global monitoring of cross-lingual worldwide news based on public-health related queries. The main monitoring module is represented over a world map with pointers to the location of origin of the news articles mentioned (location extracted from the text or publisher broad location, if the latter not available). The news articles include title, first paragraphs, publisher, publishing date/time and social media indicator (counting the number of Twitter shares).

Timeline of events based on a bar chart representing the amount of news on a particular topic over time. This permits the user to explore published news articles around events that match the search criteria.

Event categories represented by a pie chart showing the distribution of events into different categories. When the user clicks a particular segment, it displays events in the category and its percentage in the news assigned to the query.

3.3.3.2 Technology

- Frontend
    - D3.js
    - Angular
- Backend
    - Node.js
    - MongoDB
    - NewsFeed (http://newsfeed.ijs.si/) - A clean, continuous, real-time aggregated stream of semantically enriched news articles from RSS-enabled sites across the world. Uses PostgreSql for the database and provides REST API written in Python. Developed at Jozef Stefan Institute.
    - Wikifier (http://wikifier.org/) - a web service that takes a text document as input and annotates it with links to relevant Wikipedia concepts.
    - A variety of REST APIs written in Python and developed at Jozef Stefan Institute over several European projects.
- Authentication
    - Oauth2

3.3.4 MEDLINE analytics dashboard

The freely available medical/scientific research dataset MEDLINE[14] comprehends a large cover of that worldwide research (over 26 million citations) and is recognized as an important source of information in the daily life of both Public Health and Healthcare professionals. In that, QUIN made available a range of text data visualisation tools developed in the context of MIDAS. These interactive text-mining tools generating several visualisation modules, available at the MEDLINE dedicated dashboard in Pilot 1, enable the user to extract meaningful information from MEDLINE.

Currently, QUIN is focusing on the specific needs of use-case partners to develop these tools based on the existing state-of-the-art technology refocused to extract meaningful information from MEDLINE using the underlying MeSH (Medical Subject Headings) ontology-like structure. The analysis of the MEDLINE dataset is performed through WP3 and WP4 tasks. MEDLINE data is indexed with ElasticSearch and made available to analytics and visualisation tools using the Kibana technology.

The new version of the MEDLINE dedicated dashboard is based on the updated versions of ElasticSearch (6.4.1 released in September 18, 2018) and Kibana (6.4.1 released in August 18, 2018). MEDLINE articles are stored in an ElasticSearch instance hosted in premises on QUIN servers. ElasticSearch provided REST API to the modules listed below. Those modules have their own backend that uses the provided REST API.

3.3.4.1 UI Components

The MEDLINE dashboard permits the user to profit of data visualisation modules that feed on an instance of the MEDLINE open data set built in ElasticSearch. In that, Kibana is used for prototyping this tool. It also enables one to query the dataset and produce different types of data visualisation modules that can later integrate a customised dashboard, designed in agreement with the workflow of the end-user.

Prioritization of MEDLINE abstracts of articles displayed as result of a query is based on keywords. The user explores that prioritization with the movement of a pointer over the clustered concepts extracted from the articles listed, represented as word clouds. Exploration of quantitative aspects of the MEDLINE dataset through interactive data visualization modules. The latter include bar charts, scatter plots, pie charts, etc, appropriate to the data they are representing.

The elaboration of each visual module can be done by a non-technical end user directly from the structured information by a few clicks. This not only enables the

14 https://www.nlm.nih.gov/pubs/factsheets/medline.html

independence of the in-house data analysts, but also clears the availability of the IT personnel from tasks that should not be a weight in their workflow.

Moreover, from the visual modules representing aspects of interest in the MEDLINE dataset over a specific focus, the non technical user can easily create a range of monitoring dashboards to support policy-making. Furthermore, a selection of these live dashboards can be made publicly available by a given URL, excluding administrative options but maintaining its iterative nature.

MEDLINE Custom Widget (SearchPoint)

- Frontend
  - KineticJS for drawing and interacting with the widget
- Backend
  - Node.js
  - QMiner for analytics
- Authentication
  - Oauth2

Medline dedicated dashboard

The definition of this module is still in an early exploratory phase. Kibana dashboard will be used for testing and prototyping. The technology used for the development of the final version will be selected in a later phase of the project.

- Frontend
  - Kibana for prototyping
- Backend
  - ElasticSearch
- Kibana for prototyping

MedlineAtlas

- Frontend
  - D3.js
  - CSS
  - jQuery - a cross-platform JavaScript library designed to simplify the client-side scripting of HTML.
- Backend
  - QMiner (http://qminer.ijs.si/) - a data analytics platform for processing large-scale real-time streams containing structured and unstructured data. Developed in collaboration with Jozef Stefan Institute. (http://qminer.ijs.si/) - a data analytics platform for processing large-scale real-time streams containing structured and unstructured data. A product of Jozef Stefan Institute.

- Authentication
    - OAuth2

3.3.5 Stream Story

3.3.5.1 Concept

StreamStory is an exploratory data mining tool based on a system for the analysis of multivariate time series. It computes and visualizes a hierarchical Markov chain model which captures the qualitative behavior of the systems' dynamics. It provides an interactive diagram representing the cycles in the dataset, based on a sequence of states of different size (representing time spent in them) and relation between states. This includes choice of parameters, timescale or clustering algorithm to use by choice on dropdown menus. Moreover, this permits us visualisation modules, such as in Figure 3.8 representing the amount of rain in the UK as an example.

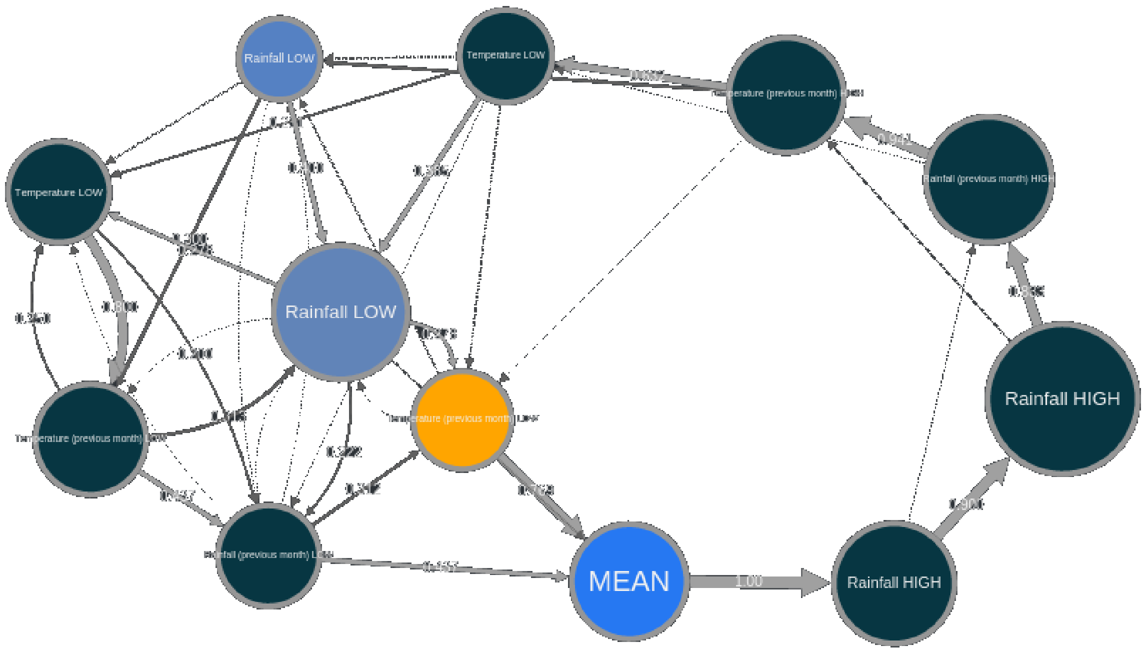

Figure 3.8: StreamStory visualization module. An illustrative example of how StreamStory represents periodicity in the data, identifying the time spent in each state and all state transitions.

This technology was not implemented in the first MIDAS prototype, as its true potential is still being explored with the MIDAS datasets that are now available. This was developed alongside the duration of the first prototype, and was improved by what was learnt from the user experiences of the first implementation. Several experiments focused the usage of the Influenzanet data set, described in D3.1, to have the results obtained adapted to other digital epidemiology topics within MIDAS. Though, due to the close analysis of priorities for the MIDAS use-case partners, StreamStory is not a priority and its implementation will not be considered on the second round of the MIDAS system, as announced in D5.1.

3.3.5.2 UI Components

3.3.5.3 Technology:

- Frontend
  - Bootstrap for design
  - jQuery to interact with the DOM
  - Cytoscape for the central graph-based visualization
  - D3 for histograms, parallel coordinates and other plots
- Backend
  - Node.js
  - QMiner for analytics
- Authentication
  - By default StreamStory uses a custom authentication module. This module can however be replaced by most standard authentication systems.

3.3.6 GYDRA Data Quality Management Tool

The initial version of the GYDRA Data Quality Management Tool was described in section 3 of the deliverable *D3.5 Data Ingestion Pilot 1* (named as Tabular data Quality assessment and Improvement, in the context of Health data, TAQIH). The tool has been renamed upon consortium recommendation to improve the previous acronym's reading relationship to negative words. GYDRA stands for Get Your Data Ready for Analysis.

The purpose of the GYDRA tool is to help non-technical people from the healthcare area in the process of exploratory data analysis (EDA). With that in mind, the tool has a user friendly visual interface. First, it provides simple yet powerful interfaces to understand the dataset, in order to gain a better understanding of the content, structure and distribution. And then, it provides data visualization and improvement utilities for different aspects regarding the data quality.

The solution described in D3.5 was focussed in datasets processing in memory. Currently, the GYDRA tool is being re-implemented (most core functionalities are ready) to support large datasets (those not fitting in memory). Figure 3.9 shows GYDRA Data Quality Management Tool's main Dashboard for an example open dataset.

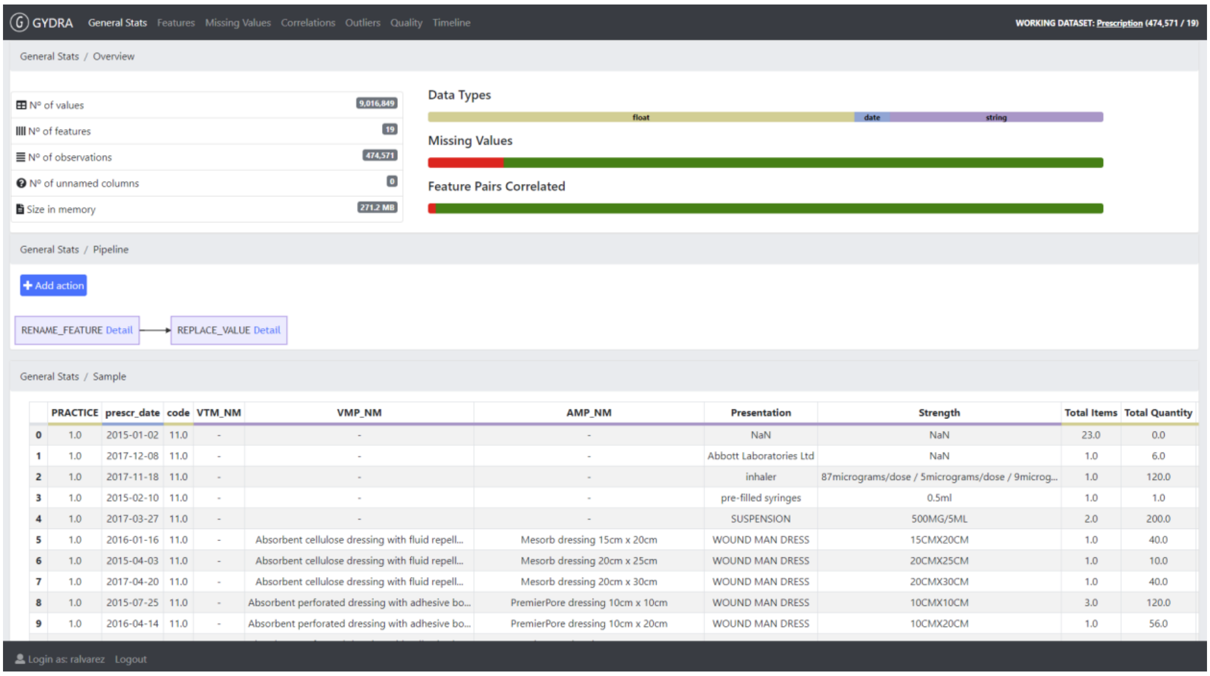

Figure 3.9: GYDRA Data Quality Management Tool's main Dashboard

3.3.6.2 UI Components

GYDRA contains a main menu bar on the top of the GUI, where items are placed from left to right following the usual iterative pipeline in EDA. First, General Stats and Features menu items provide global and detailed views of the data to gain insights about content, distribution and quality. Then, the Missing values section deals with the completeness dimension of data quality on two complementary axes, variables and samples. After that the Correlations section presents the correlations among variables, in order to help identify possible redundancies among variables or incoherent data, related to the redundancy and accuracy dimensions of data quality. Next, the Outliers section identifies outliers in the variable and samples axis which is also related to accuracy, redundancy, readability and trust dimensions in data quality. Finally, the Quality section summarizes the current state of data quality, in order to either accept the current quality of data, reject the dataset or perform a new iteration on the EDA pipeline.

In all of the tool sections a sample from the dataset is depicted as a complementary visualization, at the bottom of the GUI.

3.3.6.3 Technology

The GYDRA tool has been developed using python libraries for data pre-processing functionalities, Celery distributed task queue solution, RabbitMQ message broker software, Django python web framework (with PostgreSQL backend) and HTML5, Asynchronous JavaScript and XML (AJAX) and BootStrap responsive web development library for the user interface development.

In the MIDAS context and different sites' deployment the GYDRA tool is being deployed with Django-based frontend in the web services' devoted virtual machine, message broker and backend storage on the core services devoted virtual machine, and Celery asynchronous workers in cluster computing dedicated virtual machines. Finally the GYDRA tool is accessible through an assigned web port.

4 Implementation and Update Plan

This section describes the technology and methods which are used in the MIDAS prototypes including the current and future implementations (versions 2 and 3). The research and development is not limited to the following technology but they are the main focus on the iteration and additional technologies are possible if time and resources enable them.

4.1 Agile Development Approach with UX Testing

MIDAS Dashboard is developed in close collaboration with other MIDAS Platform components and end users. The development work is organized with Agile Sprints with Trello application based work management over all technical work packages (WP3-WP4-WP5). WP6 is following the user expectations within the UX testing which has agreed to happen both in organized test sessions and free usage of Pilot systems by selected end users. End users and developers can report and follow issues with JIRA Service Desk system which was acquired for MIDAS Project use from the Atlassian company. All MIDAS Platform codes are stored in private GitHub repositories and for this reason the technical issue tracking is handled there. The JIRA issue tracking system findings are transformed as technical issues to GitHub issue trackers. The synchronization of issue trackers is done within bi-weekly technical meetings.

4.1.1 Issue Trackers

JIRA Service Desk is a commercial service desk solution for supporting customers and teams provided by Atlassian. This software will be employed by end-users of the MIDAS platform to report issues, ask questions, or suggest features. Each end-user will be provided with the opportunity to login to the system where they can generate a request or view the status of previously generated request; users can select whether they would like to be notified of changes to the status of a request via email. There will also be an mail facility where they can use make a request via email instead of the web portal. There will be a team member from each technical work package (an agent) responsible for responding to and updating the requests. Depending on the request, these may be responded to directly, raised as an issue in GitHub against a specific technical component, or added to the Trello backlog as a new feature (which might span several components).

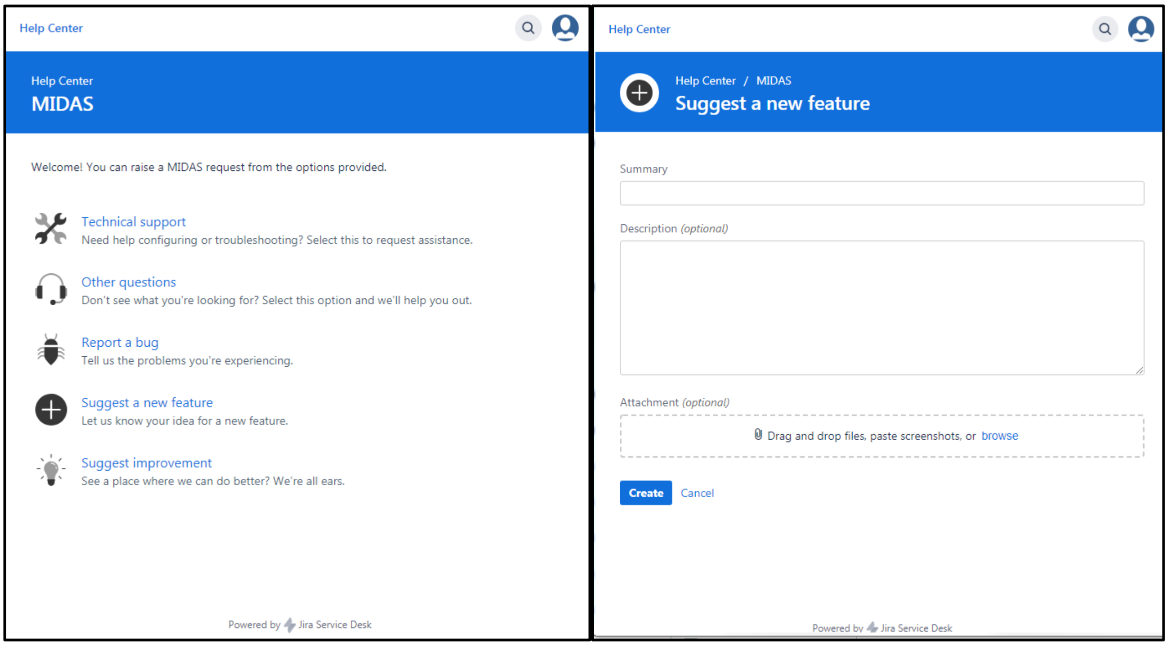

Figure 4.1 MIDAS Service Desk

Private GitHub repositories are used to manage the software being developed in the technical work packages. There are a number of repositories covering the data harmonization tool, analytics tool, and dashboard UI in addition to the backends to support these tools. Within each repository is an issue tracker which can be used to capture bugs, questions, or feature requests related to a specific component or repository. The Jira Service Desk agents from each work package will be responsible for translating Jira Service Desk requests into the appropriate GitHub issues, and monitoring the GitHub issues to update service desk requests. Due to mappings between requests and issues not being one-to-one, there is no automated process for transferring service desk requests into issues; instead this will be a manual process for the MIDAS project. When raising an issue, the agents can specify an assignee; label the issue as a bug, feature request, etc.; or set a milestone (see below Agile Schedule) and they will be notified via email when there is a change in the status of the issue (which they can feed back to the user through Jira Service Desk).

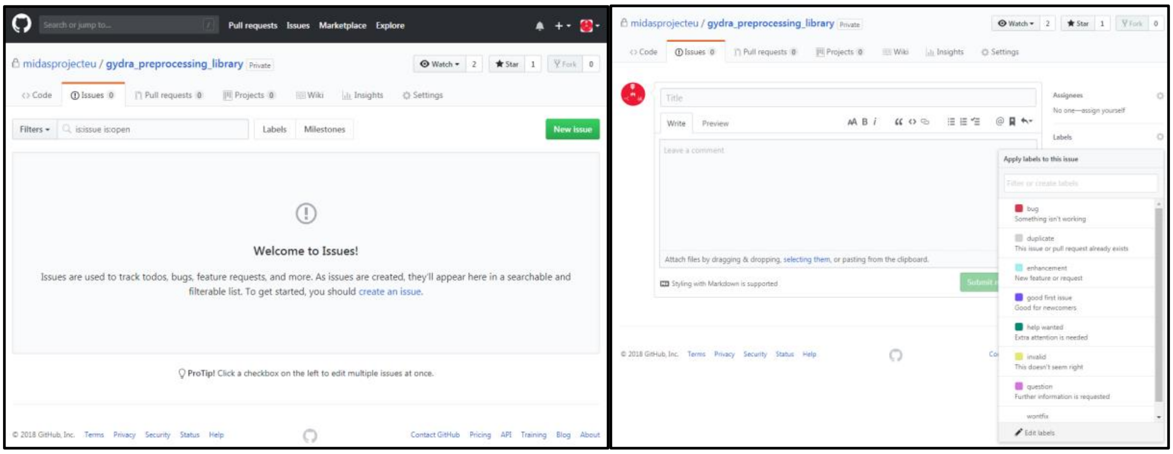

Figure 4.2 GitHub Issue Tracker

4.1.2 Trello

Trello kanban board system was agreed as an agile project management system of MIDAS technical work packages. MIDAS UI development is part of technical development workflow and UI related development steps are applied in two week sprints within all other technical development tasks. Trello enables assigning specific people to specific tasks and following the advancements of the tasks. It also enables easy updates from all signed-on users of the project in Trello.

The UI related issues tracked by JIRA service desk and transferred as technical issues in GitHub issue tracker are agreed in bi-weekly sprints and actions related to these issues are followed internally through Trello cards and reported to Issue tracing tools to show end users that their findings are taken into account. Most likely many issues reported via the issue tracker are connected to multiple work packages which is handled easily in Trello as all technical development actions are dealt in one place and UI related issues are only one part of it.

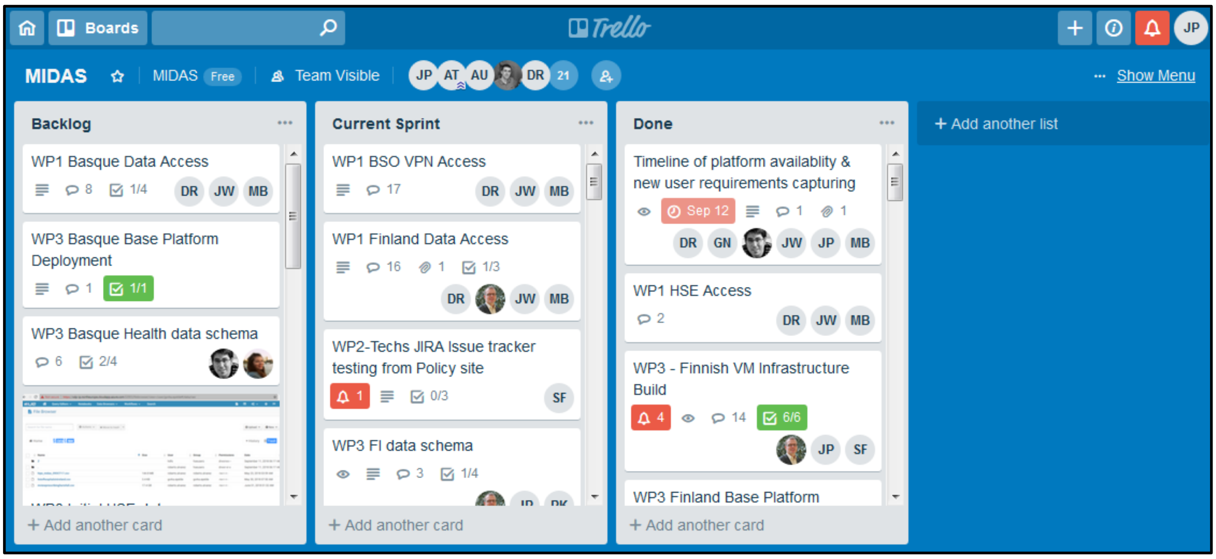

Figure 4.3 Screen capture from MIDAS Trello board

4.1.3 Agile Test Schedule

The agile test schedule was agreed together with Technical partners and Policy board members during September 2018 meeting at Dublin. The schedule is presented in Table 4.1. The schedule includes all development aspects of MIDAS Platform from technical components to user experience and impact evaluation. The Pilotwise KPIs are measured on each UI evaluation round to see how well the each development state of MIDAS Dashboard corresponds to the defined Pilot level KPIs.

Table 4.1 Agile test Schedule

AGILE TEST SCHEDULE			
Phase	**Project Month**	**Actual Date**	**Activity**
Pilot Version 1	M17	31.3.2018	Technical deliverables included integrated MIDAS platform V1 ready for deploy
	M18	30.4.2018	Deploy MIDAS platform V1 early in the month, carry out UI evaluation
	M19	31.5.2018	Bug fixes
	M20	30.6.2018	Deliverables created, final V1 released after evaluation and fixes
	M21	31.7.2018	**Work on V1**
	M22	31.8.2018	**Work on V1**
	M23	30.9.2018	**Work on V1**
	M24	31.10.2018	**FINAL V1 DATA LINKED TO BE RELEASED - LATE - HSC/ HSE / Finland (NFBC Data)**
Pilot Version 1.1 (Extra iteration)	M25	30.11.2018	**PB using system V1.0 - Start Training Feedback 1 + Tech WP leads updated user requirements (visualisation, analytics and data sources) in communication with Policy Board (directly)**
	M26	31.12.2018	V1.1 Additional Widget & Requirements from FB1, Bug Fixes,
	M27	31.1.2019	V1.1 UI evaluation
	M28	28.2.2019	Final V1.1 released after evaluation and fixes
	M29	31.3.2019	**PB using system V1.1 - Start Training Feedback 1.1 + Tech WP leads updated user requirements (visualisation, analytics and data sources) in communication with Policy Board (directly) DCU**

			Evaluation Iteration As Well
	M30	30.4.2019	V2.0- Additional Widget & Requirements from FB1.1, Bug Fixes,
	M31	31.5.2019	V2.0- Additional Widget & Requirements from FB1.1, Bug Fixes,
	M32	30.6.2019	V2.0 UI evaluation
Pilot Version 2	M33	31.7.2019	**Integrated MIDAS V2 Platform Delivered/Released PB using system V2 - Start Training Feedback 2 + Tech WP leads updated user requirements (visualisation, analytics and data sources) in communication with Policy Board (directly)**
Pilot Version 2.1 (Extra iteration)	M34	31.8.2019	V2.1- Additional Widget & Requirements from FB2, Bug Fixes,
	M35	30.9.2019	V2.1 UI evaluation
	M36	31.10.2019	Deliverables created, final V2.1 released after evaluation and fixes
	M37	30.11.2019	**PB using system V2.1 - Start Training Feedback 2.1 + Tech WP leads updated user requirements (visualisation, analytics and data sources) in communication with Policy Board (directly) DCU Evaluation Iteration As Well**
	M38	31.12.2019	Technical deliverables included integrated MIDAS platform V3 ready for deploy
	M39	31.1.2020	Deploy MIDAS platform V3 early in the month, carry out UI evaluation, bug fixes
Pilot Version 3	M40	29.2.2020	Deliverables created, final V3 released after evaluation and fixes, DCU evaluation of V3?

4.2 Implementation V2

4.2.1 MIDAS Dashboard

For the main MIDAS UI there are already identified certain functionalities which will be under investigation and research towards second iteration:

- The needed advanced visualisation methods and graphs in addition to advanced interaction within the graphs
- Methods for effectiveness assessments and feedback from the user experience testing are defined in detail based on the experiences learned with the first iteration user experience testing of MIDAS platform

With the OpenVA and GraphQL implementation the key topics which aim towards second iteration are:

- Pure GraphQL communication between Analytics Engines XDP based analytics
- GraphQL based communication with external analytics resources
- Optimized communication with analytics resources
    - Buffering of the calls
    - Bundling the calls
- Methods and implementation for automated OpenVA metadata model updates from all analytics resources.
- Robust and efficient usage of Simantics System Dynamic model resources

4.2.2 Social Media Dashboard

Known plans for future implementations:

- **Ability to dynamically control the conversation flow** - The dashboard will provide a screen where policy makers will be able to control the conversation flow. This screen will allow policy makers to have complete control over their campaigns, giving them the ability to fine grain control the questions. e.g: If a user answers 'no' to this question ask 'why' or if the user answer 'yes' to this question skip follow on questions if the user answered 'no'.
- **Stop Campaigns** – Currently, there is no way to stop a running campaign, at some point the campaign will need to be stopped as the policy makers campaign is over so the public should no longer be able to respond to that specific campaign.
- **Delete Campaigns –** Campaigns are currently kept forever and as these campaigns are using services and space that might incur a cost so the ability for Policy makers to delete campaigns is important.
- **Multi-lingual support** - We will aim to provide a solution for the chatbot to be able to talk in multiple languages. Interest has been expressed by the university of OULU (Finland) and BIOEF (Basque country) to use the chatbot in their native language.
- **Additional deeper insights** - We aim to provide deeper insights to the dashboard for future iterations, allowing policy makers to get the most out of the system.
- **Improved ability to answer questions** - We aim to improve the ability of the chatbot to have a conversation with the public, this will require improving the ability for it to answer questions from the public, as well as asking its own.

4.2.3 News Media Dashboard

Known plans for the second and third implementations details

- Concept graph representing the relationships present in the news corresponding to the query made by the user. The nodes in the graph are the most relevant concepts in the events while, the edges are displayed between the concepts that frequently co-occur in the same events.

- Clustering of events represented by a dendrogram displaying how events can be split into subgroups based on their similarity. Clicking on the leafs of the tree will display sub-clusters (to optimize the screen space).
- Integration of the MeSH classifier to enable the classification of news based on the MeSH Headings
- Representation of selected information of this dashboard through a custom widget at the main MIDAS dashboard (as discussed in 3.3.1.1).

The news custom widget was added in this D5.2 due to the need observed by use-case partners and communicated by the WP5 leader. This new custom widget is currently under development and will profit of the APIs provided by the external service Event Registry that MIDAS can access to through an acquired account.

4.2.4 MEDLINE Analytics Dashboard

Known plans for the second and third implementations details

- Representation by a MEDLINE custom widget improved from version 1, with mouseover information pop-ups and two-click redirect to article URL in PubMed.
- Automatic annotation of free text based on the MeSH Headings, loaded in a webapp text box and available through an API. This annotation based in the MeSH terms (descriptors), recurring to a classifier learning over the MEDLINE dataset, is also available for the proprietary data of free text form.

Due to the need observed in having a news custom widget together with the usefulness of the already existing MEDLINE custom widget, we understand that the mapping of MEDLINE articles in a bidimensional space announced in D5.1 is of secondary priority and will be treated as such.

During the MIDAS project we extend SearchPoint in two directions: (1) we add support for concurrent queries and (2) we add a new data adapter that queries elasticSearch to get data from MEDLINE. Originally implemented in C++, SearchPoint's components were wrapped with the Google V8 framework and exposed as a Node.js module via JavaScript API. Doing so, some code is redesigned to support concurrent memory access (e.g. elimination of smart pointers). The new data adapter is then implemented as a JavaScript callback function which queries the Elasticsearch index via REST API.

4.3 Implementation Plan for Iteration V3

In this section we have collected topics which require deeper understanding of use cases or more significant technical development work that what is possible to apply before planned V2 release.

In V1 design there was additional plans to include a server software to the MIDAS dashboard to run simulations over system dynamics model interactively within dedicated MIDAS Dashboard widget. The case was studied before V1 release but postponed due to technical difficulties. Main reasons were related to the performance issues of the open source solutions which were directly available. The models itself are developed with Vensim[15] software which is a proprietary windows software for desktop machines. As this software is not available for Linux servers an alternative approach is needed before the interactive simulator can be implemented.

Full GraphQL support for communication between OpenVA and analytics components was planned for V1, but in practice implemented only partly at V1. The current implementation is applied over traditional REST API calls which are planned to be changed to GraphQL at V3 when better communication efficiency is needed over networks.

The final integration between the MEDLINE and News open data sets through the MeSH classifier will be made available at V3, through a News dedicated dashboard where the MIDAS user will be able to explore news articles annotated with the MeSH Headings, the classification that he/she is familiar with.

5 Discussion

There are two main changes on MIDAS Dashboard development from V1 to V2. The first is the agile development approach which focus was improved during the summer. It enables MIDAS Platform to development to take into account the end-user requirements and thus deliver better solution for each pilot. The second change is the pilotwise deployments of the system. These will add pilot specific analytics to be supported by MIDAS Dashboard. These analytics are different on each pilot which tests well the flexibility of the MIDAS Platform. The Pilotwise testing enables us to measure our success to fulfill the KPIs of the Pilot sites defined at D5.1. These measures are reported with deliverable D5.4 and D5.5 Visual analytics tool(s) - MIDAS Dashboard functional prototype V2 and V3 respectively.

The changes in MIDAS UI from the first iteration by the UX testing have improved the MIDAS Dashboard significantly and the new wizard approach from visualisation to data will enable more straight forward process to create the new widgets. The new authentication approach to support the single sign-on is more challenging to implement than earlier approaches but has significant effect on user experience when authentication requests between different end user views disappear. One significant change in V2 development is the direct feedback from the end-users through JIRA Service Desk. It brings the end-users closer to developers and helps end-users to

[15] http://vensim.com/

follow the development at the same time it helps developers to understand the hands on usage of the system by end users.

6 Conclusion

The MIDAS Dashboard V2 will deliver the first Pilot level experience to end-users with their own data and pilotwise dedicated analytics. The agile development with agreed testing schedule enables technical partners to develop MIDAS Platform to deliver solution which corresponds to the real needs of the end users of each Pilot.

7 Appendix 1 OpenVA Data Model

Figure A1.1 shows the data model between OpenVA and UI in MIDAS Dashboard.

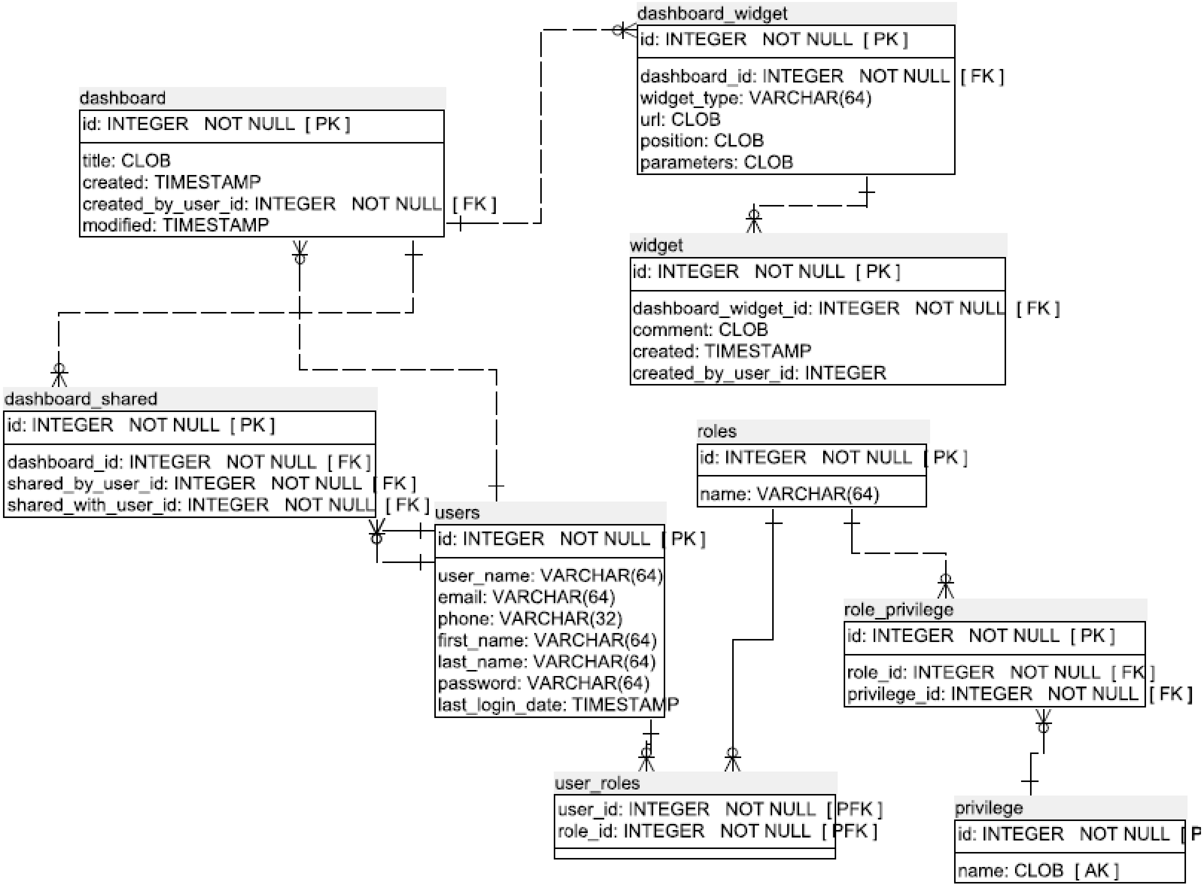

Figure A1.1: Data model of OpenVA

8 Appendix 2 Updated Requirements for Custom Widget

Inputs: ID of a Div (DOM element)
Render any custom visualization within the container Div. Any web-based technology that is supported by all major browsers e.g. HTML Canvas, SVG, D3 etc. can be used to render the visualization

Output:
The plugin renders the required visualization within the div (input div) and handles required interactions e.g. zoom, pan, data point selection etc. Since the widgets are resizeable and responsive therefore, the rendered visualization needs to be responsive as well.

Data Model:
A generic data model is used to render all the visualizations in the MIDAS dashboard.

Custom widget/ plugin should be accessible easily. For example:
customPlugin.renderVisualization("graphDiv1", dataModel);

```
var dataModel = {
        chartStaticImgUrl: "",
        chartTitle: "",
        chartSubTitle: "",
        xLabel: "",
        yLabel: "",
        valueRange: null,
        timeUnit: null,
        startDate: null,
        endDate: null,
        chartData: {
           xAxisType: "",
           yAxisType: "",
           xData: [],
           yData: []
        }
    };
```

Field	Data Type	Definition
chartStaticImgUrl (optional)	string	url of the generated static image.
chartTitle	string	Main title of the graph

chartSubTitle (optional)	string	string of object of interest names separated by comma,
xLabel	string	Label/title that defines the X-Axis.
yLabel	string	Label/title that defines the Y-Axis.
valueRange (optional)	JSON Array	Contains **min** and **max** values for Y-axis if required. for example [0, 100]
threshold (optional)	integer/float	To draw a target/goal/threshold line at a specified Y-Axis value.
timeUnit (optional)	string	one of: “sec, “min”,”hour”,”day”,year (if required)
startDate (optional)	DateTime	ISO JavaScript date
endDate (optional)	DateTime	ISO JavaScript date
chartData	JSON Object	Contains all the required data to draw a graph. Explanation below.

chartData:

Field	Data Type	Definition
xAxisType	string	To calculate X-axis scale so that functionalities like zooming, panning and Inter-graph synchronization can be achieved. Options: 1. **value:** Numerical axis, suitable for continuous data. 2. **category**: Category axis, suitable for discrete category data. 3.**time:** Time axis, suitable for continuous time series data.
yAxisType	string	To correctly calculate Y-axis scale so that functionalities like zooming, panning and Inter-graph synchronization can be achieved. Options: 1. **value:** Numerical axis, suitable for continuous data.

		2. **category**: Category axis, suitable for discrete category data. 3.**time:** Time axis, suitable for continuous time series data.
xAxisDateFormat	string	Date format for the x-Axis labels. For example, YYYY-MM or YYYY-DD-MM or YYYY or other combination.
xData	JSON Array	Array of values for X-Axis
yData	JSON Array	It is an array of arrays. Explanation below

yData:

Field	Data Type	Definition
yData [0]	JSON Array of Objects	Array on zero index defines the type of visualization to be drawn on Y-axis. for example: [{ "seriesType":"line",label":"%"} { "seriesType":"line", “label":"mean"}] **SeriesType**: defines the type of particular series. one of the values defined in https://datavizcatalogue.com/ For example, linegraph, scatterplot, barchart, piechart, heatmap etc. **label**: defines a label for series used for the legend and data point tooltips.
yData[1…]	JSON Array of series data	Each subsequent array in the yData array, represents data of defined series. So according to above example first array (yData[1]) will contain data for “%” series and second array (yData[2]) contains data for “mean” series. Make sure the order of data in these arrays should be according to the order of data in xData array.

Visualization support with current data model:

The above data model supports following visualizations. It also supports combining of multiple visualizations into one graph widget. For example, combine line and scatter plot visualizations in one graph object.

1. Line graph
2. Scatterplot
3. Bar chart
4. Horizontal bar chart,
5. Stacked bar chart
6. Horizontal stacked bar chart
7. Pie chart

8. Radar chart
9. Bubble chart
10. Candle chart
11. Area chart
12. Timeline
13. Correlation matrix.
14. Map chart